International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 4 ǁ April 2023 ǁ PP. 612-616

A Continuous Real-Time Action Detection Approach to Sign Language

Yashwini Chaudhary[*1], Tanishq Gupta[*1]
*[*1]Department of Computer Science and Engineering , SRM Institute of Science and Technology,U.P., India*

Abstract
One use of Artificial Intelligence and Machine Learning is the detection of sign language, which has grown in importance and effectiveness for humans. Research in this area is now ongoing. In earlier efforts, static sign detection was accomplished with the aid of a straightforward Convolutional Neural Network powered by deep learning. In order to identify the action taken by the user, this approach is based on continuous real-time action detection of visual frames. After locating critical spots utilizing mediapipe holistic, which contains facial, position, and hand information, the model uses an LSTM neural network model. The data is pre-processed, labels and features are created, and critical value points for raining and testing are gathered as part of the proposed effort. The weights are saved, and the model is assessed using the correctness of the confusion matrix. This study uses continuous real-time image frame detection to identify sign language.

Keywords: *Artificial Intelligence,Deep Learning, LSTM, Machine Learning, Mediapipe Holistic,Sign Language, Real-Time Image Frame Detection*

Date of Submission: 11-04-2023 Date of acceptance: 25-04-2023

I. INTRODUCTION

Human beings with hearing disability and/or speech incapacity depend completely on signal language, a non-verbal language, to speak with the ones round them. It combines a diffusion of nonverbal conversation techniques, which include finger spelling, hand gestures, frame language and facial feelings. As opposed to spoken phrases, sign languages (on occasion called signed languages) use the visible-guide modality to speak, which means that. Linguists classify spoken and signed communication as types of herbal language. That means that each superior via time, with careful planning and through an precis, protracted developing old manner. The maximum not unusual shape of human expression or conversation is speech, which need to be replaced at the same time as speech is impaired with the resource of a kinesthetic method of conversation. Sign Language, additionally, known as visual language, commonly has five vital characteristics, along with hand form, orientation, movement, role and extras like mouth form and brow movement. But, due to the fact there's fallacious conversation among them, people who do not understand signal language usually devalue or reject people with such an impairment.It involves recording someone making hand gestures, reading the video, and sending it to the caution model, which then predicts terms one after the opposite. A few researchers contend that early human beings used signs and symptoms and signs and symptoms to talk long in advance than the improvement of spoken language. If you have raised your hand to call a cab, driven your index finger toward your lips to calm a rowdy teenager, or pointed to an item on the menu, you have used sign language in its maximum essential form. We are capable of speaking with people who have listening to loss or linguistic disabilities via the use of signal language. A spread of gestures performed with the fingers, fingers, arms, head, and additionally, facial expressions that useful resource inside the communique of the deaf and tough of taking note of with the ones round them. An alphabet is represented by using every of the 26 hand gestures that make up the unwritten language. These networks can display a selection of characteristics counting on the amount and form of layers used, making them generally useful for an expansion of jobs.

1.1 Problem Statement

In our progressive society, it's necessary to fraternize with people whether for recreation or a purpose. Communication is important for every human being. However, people who have a hearing disability and/ or a speech disability need a different way to communicate other than verbal communication. They resort to sign language to communicate. However, Sign Language requires a lot of training to be understood and learned, and not every person may understand what sign language gestures mean. Learning sign language is also time-consuming as there are no effective tools for recognizing sign language. Hearing or speech impaired people who know Sign Language need a translator to explain their thoughts to other people in an effective manner. To help overcome these problems, this project aims to bridge the gap.

1.2 Objectives

1. To generate large amounts of appropriate datasets using video classification.
2. Take the captured videos, break them down into frames of images that can then be passed onto the system for further analysis and interpretation.
3. To apply appropriate frame pre-processing techniques in order to remove the noise and obtain the ROI.
4. To design the model and architecture for RNN and LSTM to train the pre-processed video frames and achieve the maximum possible accuracy.
5. To develop an algorithm to prognosticate the gesture in real time.

II. LITERATURE SURVEY

REAL TIME FINGER TRACKING AND CONTOUR DETECTION FOR GESTURE RECOGNITION USING OPENCV [1]: According to Gurav and Kadbe, only clear and steady signs can be detected and the maximum speed for these simple signs is 30 fps.(note this is with some distinct computer properties and the use of only one hand). Furthermore, this only works with a plain contrasted background and with the hand kept close to the camera.

COUPLED HMM-BASED MULTI-SENSOR DATA FUSION FOR SIGN LANGUAGE RECOGNITION [2]: The authors try different hand tracking approaches to subsequently prognosticate 25 different signs with a HMM.To track the hands three approaches were used with two detectors Kinect, leap motion and a combination of both attaining the conclusion that the combination of both(Kinect and leap motion) works more. The database used in the paper corresponded of 2000 word samples from 25 different Indian sign language gestures. Some of the samples are the gesture of the word and some are made by spelling the letters of the word. The maximum precision reached is 90 percent. When sign language detection models are used, the scalability challenge arises because the number of HMMs that need to be trained boosts with the increased vocabulary.

REAL-TIME SIGN LANGUAGE GESTURE (WORD) RECOGNITION FROM VIDEO SEQUENCES USING CNN AND RNN [3]: Masood et al(2018) [3] applied CNN to detect the position of the hands and RNN to detect patterns of the signs through time. In this paper, the authors achieve the recognition of 46 signs with a precision of 95.2%. The results are certifiably all right but are only possible with some approaches that rule out the possibility of real time recognition. In order to detect the hands, the authors remove all the background so only the hands appear in the image. The quality of the camera used by the authors isn't specified.

A MODIFIED - LSTM MODEL FOR CONTINUOUS SIGN LANGUAGE RECOGNITION USING LEAP MOTION [4]: In this paper, the authors[4] have offered a unprecedented framework for uninterrupted-SLR employing the Leap motion detector. A modified LSTM framework has also been proposed for the recognition of sign words and sentences. Average precision of 72.3 percent and 89.5 percent have been recorded on the signed sentences and isolated sign words, respectively. The recognition performance can be enhanced by expanding anothis training data for better model learning.

A SIGNER INDEPENDENT SIGN LANGUAGE RECOGNITION WITH CO-ARTICULATION ELIMINATION FROM LIVE VIDEOS: AN INDIAN SCENARIO [5]: The system produced by the authors[5] recognizes ISL gestures from mobile camera videos with no additional detectors to detect hand regions. Single- handed static and dynamic gestures and double- handed static gestures are identified in this system. The system uses ROI algorithm for characteristic extraction. After pre-processing the image, the centroid of the double image is estimated. The static and dynamic gestures are identified using the gesture separation method. The advantage of this system is that it's provident and can be executed with a mobile camera, making it truly user-friendly. But the disadvantage is that it isn't effective under cluttered backgrounds and different illumination conditions.

A DEPTH-BASED INDIAN SIGN LANGUAGE RECOGNITION USING MICROSOFT KINECT [6]: Overlapping signs, double hand signs, and signs unique to ISL were recognized successfully using this system[6]. The advantage of this design is that the average recognition precision was enhanced up to 71.85 percent with this system. The system achieved 100 percent precision for a many of the signs but the system doesn't consider the surroundings of gestures, leading to incorrect translations on multiple occasions.

VIDEO-BASED ISOLATED HAND SIGN LANGUAGE RECOGNITION USING A DEEP CASCADED MODEL [7] : The authors[7] used a cascaded framework of SSD, CNN, and LSTM from RGB video recordings to propose a deep- based model for systematic hand sign recognition. The preciseness and complexity of hand sign recognition were enhanced by this model. In case of an uncontrolled surround such as swift hand movements, it delivered fast processing. Using additional data, the precision of detection can be enhanced.

AN EFFICIENT BINARIZED NEURAL NETWORK FOR RECOGNIZING TWO HANDS INDIAN SIGN LANGUAGE GESTURES IN REALTIME ENVIRONMENT [8]: Considering the challenges of sign language recognition, on targeted embedded platforms, authors[8] have proposed the unprecedented framework of a binarized neural network with binary values of weights and activations utilizing bitwise operations. The advantage is using this framework achieves an overall precision of 98.8 percent which is more advanced than othis existing methodologies while the disadvantage is this system misclassifies some gestures of M, N, E because of their resemblant kind of configurations, and also, the proposed BNN framework is limited with small number of classes of gestures.

III. SYSTEM ANALYSIS

3.1 Issues in the existing system

The field of sign language detection is fraught with difficulties and issues. There are several sign languages used in various nations, and even within a single nation, regional differences may exist. It is challenging to create a universal sign language recognition system because of the lack of standards.

Systems for detecting sign language need to be trained on a lot of data. To record and annotate a large number of sign language users, it might be challenging to gather such data. This can be difficult, particularly in nations where sign language is not often used. Hand gestures, face expressions, and body motions are all used in sign language. These motions can be intricate and challenging to identify, particularly in real-world situations with imperfect lighting and backgrounds. Computer vision and machine learning techniques, which are utilized for sign language detection, are still in their infancy and have significant limitations. This may affect the precision and dependability of systems that detect sign language. Despite these obstacles, scientists and programmers are attempting to create sign language recognition systems that are precise, dependable, and usable by sign language users.

3.2 Proposed System

Using long Short-Term Memory networks and deep learning methods is one suggested approach to overcoming the difficulties in sign language identification. This method can increase the accuracy and dependability of existing sign language recognition systems by addressing some of their major constraints**.** Deep learning techniques allow computers that recognize sign language to learn to identify patterns in sign language motions and facial expressions, even when such patterns change between various areas and nations. This can aid in overcoming sign language's lack of uniformity. Large volumes of data, including sign language motions recorded from various sources, may be used to train deep learning systems. This might assist us get over the lack of annotated sign language data.

For identifying intricate sign language gesture sequences, LSTM networks are an excellent choice. Even when done in real-world circumstances with varied lighting and background conditions, these networks can learn to detect patterns in sign language motions. The ambiguity in sign language recognition may be reduced by teaching LSTM networks on a significant quantity of sign language data and teaching them to understand the context and meaning behind gestures. Deep learning algorithms have shown potential in enhancing sign language recognition systems' precision and dependability. The complex, sequential characters of sign language motions and facial expressions can be more effectively handled by LSTM networks in sign language recognition systems.In conclusion, employing LSTM networks and deep learning methods may be a potential answer to the problems associated with sign language identification. These strategies can aid in enhancing the precision and dependability of sign language recognition systems, hence enhancing their usability for sign language users.

IV. METHODOLOGIES

4.1 Image Processing And Computer Vision

The OpenCV module in Python supports a variety of image processing and computer vision techniques, such as object detection, feature extraction, and image segmentation. These techniques can be used to detect hand or body motions in videos or images of sign language. By integrating image processing techniques and machine learning algorithms, OpenCV can be used for sign language detection and recognition. The initial stage in detecting sign language is to capture an image or video of anyone performing sign language. This can be achieved with a camera or a webcam. To detect the face, hand and body region in an image or video stream, OpenCV is deployed. Once the region has been identified, its movement and form and orientation can be tracked and analysed over time.

4.2 Machine Learning

Python's scikit-learn library provides a wide range of machine learning algorithms, such as support vector machines, decision trees, and random forests. These algorithms can be used to train a model to recognize sign language gestures based on extracted features.

4.3 Deep Learning

TensorFlow is a Python package that has extensive tools for constructing and training deep neural networks. These networks can be utilised for sign language detection and segmentation such image categorization, object detection, and sequence modelling. TensorFlow provides a number of algorithms for sign language detection depending on the topic and dataset. We employed the LSTM method in this research. LSTMs are RNNs designed to handle long-term dependencies in sequential data. Memory cells are utilised to retain and retrieve information over a long period of time. LSTMs are often used for detection of sign language tasks that require sophisticated hand movements and temporal patterns. Generally, TensorFlow offers a diverse variety of approaches for sign language detection, depending on the application and dataset. To achieve the greatest performance, it is critical to select the most suited approach for the task and to experiment with various models and hyper-parameters.

4.4 Evaluation and Accuracy

The confusion matrix and accuracy are two commonly used measures for evaluating the performance of sign language detection models.A confusion matrix is a table that compares predicted and observed labels for a classification problem. There are four entries in the table: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (FN). A true positive indicates that the model correctly identified a sign language gesture, a true negative indicates that the model correctly identified a non-gesture, a false positive indicates that the model incorrectly identified a non-gesture as a gesture, and a false negative indicates that the model incorrectly identified a gesture as a non-gesture.

Accuracy is a metric that evaluates the model's accuracy as a percentage of all predictions. It is determined by dividing the total number of forecasts by the number of right predictions (i.e., the sum of TP and TN) (i.e. the sum of TP, TN, FP, and FN).To assess the effectiveness of a sign language detection model, we can compute the confusion matrix and accuracy on a validation or test set. The confusion matrix can help us understand the types of errors made by the model, such as whether it frequently misclassified non-gestures as gestures.Although accuracy can offer an overall measure of model performance, it should be used in conjunction with other metrics such as precision, recall, and F1-score to gain a more complete understanding of the model's strengths and limitations.

Overall, the confusion matrix and accuracy are two significant metrics that may be used to evaluate the performance of sign language identification models, and they should be calculated on a validation or test set to confirm that the model generalises well to new data.

V. EXPERIMENTS AND RESULTS

5.1 Dataset Generation

We have collected the keypoint values for training and testing our media-pipe model using OpenCV to record ourselves while performing different Sign Language Gestures. The collection should contain a wide diversity of angles for each sign. The video data has to be preprocessed in order to extract the important properties. This includes converting the video frames into a representation that models can easily understand, such as a collection of 2D images or a stream of motion vectors. The data must be preprocessed after collection in order to be ready for LSTM model training.

5.2 Gesture Classification

Data preprocessing and labeling are steps taken to clean and transform raw data into a format suitable for analysis. Preprocessing ensures that the input data is in a consistent and usable format, while labeling involves identifying and classifying the different signs in the input video. Feature extraction is the process of selecting relevant features from the input data that are essential for classification. By selecting relevant features, the model can learn to differentiate between different signs and improve its accuracy.

5.3 Testing

The trained sign language identification model is deployed in a real-world situation during real-time testing. In the instance of sign language identification, this entails recording a live video feed of someone signing and using the trained model to predict the sign language words or phrases being signed.You may use OpenCV to grab the video stream from a camera or a video file to do real-time testing. To retrieve the preprocessed keypoint data, repeat the keypoint extraction and preprocessing processes used during the training phase. Finally, the preprocessed data may be fed into the trained LSTM model to generate the predicted sign language words or phrases. Real-time testing can be difficult since the model must process the video stream in real-time, which can be computationally demanding. Techniques such as batch processing and parallelization can be used to increase performance.

VI. CONCLUSION

In this report, a functional real-time video based sign language recognition for Deaf and dumb people has been developed . The LSTM model architecture is well-suited for modelling sequential data such as sign language motions and may be trained using Mediapipe Holistics' preprocessed keypoint data. The model's ultimate accuracy score of 0.8 suggests that it can predict sign language words or phrases for a range of sign language motions. However, there is always room for development in terms of accuracy, particularly for more sophisticated sign language movements or settings with changing lighting or camera angles. Furthermore, the model may need to be modified for usage with other sign languages or dialects, which may need extra data collecting and preprocessing work.

REFERENCES

[1]. Ruchi Manish Gurav and Premanand K. Kadbe. “Real time finger tracking and contour detection for gesture recognition using OpenCV”. In: 2015 International Conference on Industrial Instrumentation and Control, ICIC 2015. Institute of Electrical and Electronics Engineers Inc., July 2015,pp.974–977.ISBN:9781479971657.DOI:10.1109/IIC.2015.7150886.

[2]. Pradeep Kumar, Himaanshu Gauba, Partha Pratim Roy, and Debi Prosad Dogra. “Coupled HMM-based multi-sensor data fusion for sign language recognition”. In: *Pattern Recognition Letters* 86 (2017), pp. 1–8.

[3]. Sarfaraz Masood, Adhyan Srivastava, Harish Chandra Thuwal, and Musheer Ahmad. “Real-Time Sign Language Gesture (Word) Recognition from Video Sequences Using CNN and RNN”. In: *Intelligent Engineering Informatics*. Ed. by Vikrant Bhateja, Carlos A. Coello Coello, Suresh Chandra Satapathy, and Prasant Kumar Pattnaik. Singapore: Springer Singapore, 2018, pp. 623–632. ISBN: 978-981-10-7566-7.

[4]. A. Mittal, P. Kumar, P. R. Roy, R. Balasubramanian, B. B. Chaudhuri, “A Modified-LSTM Model for Continuous Sign Language Recognition using Leap motion”,2019. https://ieeexplore.ieee.org/abstract/document/8684245

[5]. Athira, P.K. & C J, Sruthi & Lijiya, A.. (2019). A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario. Journal of King Saud University - Computer and Information Sciences. 34. 10.1016/j.jksuci.2019.05.002.

[6]. T. Raghuveera, R. Deepthi, R. Mangalashri, R. Akshaya, “A depth-based Indian Sign Language recognition using Microsoft Kinect”, 2020. https://link.springer.com/article/10.1007/s12046-019-1250-6.

[7]. R. Rastgoo, K. Kiani, S. Escalera, “Video-based isolated hand sign language recognition using a deep cascaded model”, 2020. https://link.springer.com/article/10.1007%2Fs11042-020-09048-5

[8]. M. Jaiswal, V. Sharma, A. Sharma, S. Saini, R. Tomar, “An Efficient Binarized Neural Network for Recognizing Two Hands Indian Sign Language Gestures in Real-time Environment”, 2020 https://ieeexplore.ieee.org/ abstract/document/9342454