2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)

Real-time Gesture Based Sign Language Recognition System

[1]Jeet Debnath
PG Scholar
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
jeetdeb232@gmail.com

[2]Praveen Joe I R
Assistant Professor Senior Grade II
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
praveen.joe@vit.ac.in

2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) | 979-8-3503-6482-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ADICS58448.2024.10533518

***Abstract*— Sign language is a vital mode of communication for the deaf and dumb community. This research presents a robust and real-time Gesture-Based Sign Language Detection System that leverages computer vision and deep learning techniques. The system is designed to recognize and interpret American Sign Language (ASL) gestures, enabling efficient communication between individuals who are proficient in ASL and those who are not. The core of the system utilizes Python, OpenCV (Open-Source Computer Vision Library), and MediaPipe Holistic for real-time hand and body pose estimation. By accurately tracking the movements of hands and key body parts, the system captures the nuances of sign language gestures. The captured data is then fed into a Long Short-Term Memory (LSTM) neural network, which excels in sequence modeling tasks. The LSTM model is trained on a comprehensive dataset of ASL gestures, encompassing a wide range of signs and expressions. Transfer learning techniques are also employed to fine-tune the model for improved performance in sign language recognition. The model's architecture allows it to learn the temporal dependencies and context inherent in sign language, making it capable of recognizing gestures within sentences or phrases. The system's evaluation demonstrates its effectiveness in real-world scenarios, achieving high accuracy and low latency in sign language recognition. It opens new avenues for accessible and inclusive communication, aiding both deaf and hearing individuals in bridging the communication gap. Future work may explore the integration of natural language processing (NLP) to facilitate two-way communication between sign and spoken language.**

In conclusion, this Gesture-Based Sign Language Detection System represents a significant step towards harnessing the power of computer vision and deep learning to make sign language more accessible and inclusive in various domains, including education, accessibility, and social interaction.

Keywords*— *Gesture Recognition, Sign Language Detection, LSTM

I. INTRODUCTION

Sign language, a complex and expressive mode of communication predominantly used by the deaf and hard-of-hearing community, plays a crucial role in fostering effective interpersonal connections and enabling individuals to access information, education, and various aspects of daily life. With millions of people worldwide relying on sign language as their primary means of communication, the development of technologies that bridge the gap between sign language users and the broader community is of paramount importance. In this context, we present a groundbreaking research endeavor: the "Gesture-Based Sign Language Detection System."

The "Gesture-Based Sign Language Detection System" represents a multidisciplinary fusion of CV, ML, and deep learning techniques, coordinated to enable smooth communication between individuals who use sign language and those who might not be fluent in it. At its core, this system leverages the power of Python, OpenCV, MediaPipe Holistic, and Long Short-Term Memory (LSTM) neural networks to accurately and efficiently interpret American Sign Language (ASL) gestures.

The Significance of Sign Language

Sign language, as a form of non-verbal communication that relies on hand shapes, facial expressions, and body movements, is not merely a set of gestures but a rich and nuanced language in its own right. ASL, for instance, boasts a lexicon and grammatical structure that rival spoken languages. However, while it is a vital tool for many, The linguistic divide between those who use sign language and others who do not can lead to social, educational, and employment challenges for the deaf community.

The Role of Technology

The potential of technology to bridge this communication gap has long been recognized. Early efforts in this direction involved the development of sign language recognition systems based on primitive hand-tracking techniques. However, these systems often struggled to capture the intricacies and fluidity of sign language, limiting their practical utility.

Recent advancements in computer vision and deep learning have breathed new life into the pursuit of effective sign language recognition. The advent of MediaPipe Holistic, a holistic human pose estimation pipeline developed by Google, has enabled the simultaneous tracking of hand and body movements, bringing a new level of accuracy to gesture recognition. Additionally, the application of LSTM neural networks, renowned for their ability to model sequential data, has unlocked the potential for recognizing not just individual signs but entire sentences or phrases in sign language.

Research Objectives

The primary objective of this paper is to introduce and elucidate the architecture and capabilities of the Gesture-Based Sign Language Detection System. This system, grounded in state-of-the-art technology, seeks to address the following key goals:

Real-time Sign Language Interpretation: The system endeavors to provide real-time interpretation of ASL gestures,

enabling spontaneous and natural communication between sign language users and non-signers.

Accuracy and Robustness: The system aims to achieve high accuracy in recognizing a large variety of sign language movements, including variations in handshapes, movements, and facial expressions. Robustness in diverse lighting and environmental conditions is a critical aspect.

Structure of the Research Paper

This research paper is structured as follows: In subsequent sections, we delve into the methodology, dataset, and training procedures, providing insights into the technical aspects of the Gesture-Based Sign Language Detection System. We then present the results of our experiments, demonstrating the system's performance in real-world scenarios. Finally, we discuss potential future directions and applications, underscoring the transformative potential of this technology.

In essence, the "Gesture-Based Sign Language Detection System" represents a significant stride towards harnessing cutting-edge technology for social good, enabling effective communication and fostering inclusivity for the deaf community.

II. Related Works

[1] This paper focuses on various methods of interpretation systems for deaf-mute communication. It categorizes these communication methods into two main groups: Portable Communication Devices and Online Learning Systems.

(i) Portable Communication Devices: This category includes wearable communication methods that aim to facilitate communication for deaf individuals. These methods are designed to be used on the go and offer real-time communication support. The three subcategories within this group are:

- Glove-based Methods: These methods likely involve wearable gloves equipped with sensors or technology to interpret sign language or gestures made by the user. The gloves translate these movements into text or speech, enabling communication with non-deaf individuals.
- Keypad Methods: Keypad-based methods may involve specialized devices with keypads that allow deaf users to input text or symbols, which are then displayed or vocalized to facilitate communication.
- Handy Cam Touch-screen Methods: This category may involve using handheld cameras or touchscreen devices to capture sign language or gestures and translate them into text or speech for communication purposes.

(ii) Online Learning Systems: This method represents an alternative approach to overcome the need for external devices. Online learning systems likely involve digital platforms or applications that teach sign language or communication skills to both deaf and non-deaf individuals. These systems can enable effective communication without the need for specialized hardware.

[2] The Indian Sign Language Recognition (ISLR) system described in this study uses two essential modules for pattern recognition: feature extraction and classification. Here's a brief literature survey based on the information provided:

Feature Extraction: The first module in the proposed ISLR system is feature extraction. It utilizes the Discrete Wavelet Transform (DWT) as the basis for extracting features from sign language data. DWT is a well-established technique in signal processing and image analysis. The use of DWT suggests that the paper acknowledges the importance of capturing both time and frequency information, which is often crucial.

Classification: In the second module, which is called classification, sign language motions are recognized using the features that were extracted. The paper mentions the use of the nearest neighbor classifier. Nearest neighbor algorithms are commonly used in pattern recognition tasks due to their simplicity and effectiveness. The choice of classifier can significantly impact the performance of a recognition system.

Experimental Results: The paper reports experimental results, indicating that the proposed ISLR system achieves a remarkable maximum classification accuracy of 99.23%. This is a crucial finding as it demonstrates the effectiveness of the DWT-based feature extraction method combined with the nearest neighbor classifier in accurately recognizing Indian Sign Language gestures. Furthermore, the use of the cosine distance classifier is mentioned, indicating that different classifiers were evaluated during the experimentation.

[3] The paper proposes a method for hand gesture recognition that involves several steps:

Hand Region Segmentation: Initially, the hand region is segmented from the background. This is achieved by applying a skin color model in the YCbCr color space. YCbCr is a color space that separates luminance (Y) and chrominance (Cb and Cr) components. Skin color detection in this color space can help isolate the hand region from the rest of the image.

Thresholding: The foreground (hand area) and backdrop are separated in the following step by applying thresholding. Thresholding is a common technique in image processing where a certain pixel value threshold is used to classify pixels as foreground or background based on their intensity or color.

Template-Based Matching with PCA: Finally, the paper uses a template-based matching technique for hand gesture recognition. Principal Component Analysis (PCA) is employed for dimensionality reduction. PCA is a method that reduces the dimensionality of data while preserving as much of the variability in the data as possible. In the context of this paper, it is likely used to extract essential features from the segmented hand region.

The PCA-transformed features are then compared to templates or reference patterns representing various hand gestures. This comparison allows the system to recognize and classify the hand gestures based on the similarity between the transformed features and the templates.

This approach combines skin color-based segmentation, thresholding, and PCA-based dimensionality reduction for hand gesture recognition. By using PCA, the authors can reduce the complexity of the feature space while maintaining

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

discriminative information, making it suitable for real-time gesture recognition applications.

[4] This paper proposes a method for hand gesture recognition that involves several steps:

Hand Gesture Recognition for Communication: Begin by discussing the importance of hand gesture recognition systems for individuals with speech and hearing disabilities. Explain how such systems can enable effective communication.

Previous Approaches: Review existing methods and approaches for hand gesture recognition using image processing. This could include techniques based on computer vision, machine learning, and deep learning. Discuss the strengths and limitations of these approaches.

Gesture Datasets: Mention any publicly available gesture datasets that have been used in previous research. Highlight the significance of having diverse and well-labeled datasets for training and evaluating gesture recognition systems.

Image Processing Techniques: Describe common image processing techniques that are typically used in hand gesture recognition, such as background subtraction, skin color detection, contour analysis, and feature extraction. Discuss how these techniques contribute to accurate gesture recognition.

Machine Learning and Deep Learning Models: Explore the use of machine learning and deep learning models in gesture recognition. This could include techniques like Support Vector Machines (SVM), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). Discuss the advantages and challenges of using these models.

Real-time Recognition: Emphasize the importance of real-time gesture recognition, especially for practical applications. Discuss any previous work that has focused on achieving low-latency recognition.

Challenges and Open Problems: Highlight the challenges and open problems in the field of hand gesture recognition. This could include issues related to occlusion, lighting conditions, and the need for robustness in real-world scenarios.

Applications: Discuss the various applications of hand gesture recognition systems beyond communication, such as in human-computer interaction, virtual reality, and gaming.

Comparative Analysis: If applicable, provide a comparative analysis of different hand gesture recognition systems in terms of accuracy, speed, and usability.

III. Methodology

A. Data Collection

To develop an effective Sign Language Detection System for my master's thesis, a meticulously crafted dataset of sign language gestures was collected (see Fig. 1). The dataset was organized to include 30 sequences for each sign language gesture, with each sequence comprising 30 frames. This structured dataset provided a rich and diverse representation of sign language gestures, enabling a comprehensive training and testing environment.

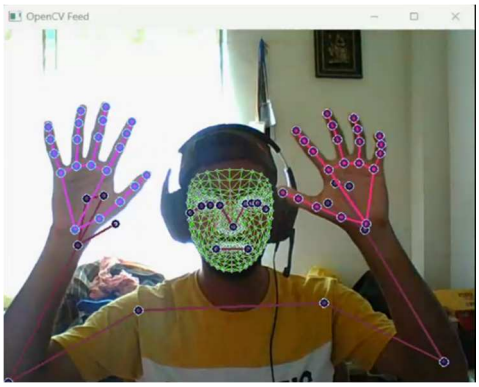

Fig. 1. Data Collection

B. Data Preprocessing

The collected dataset underwent thorough preprocessing to optimize its suitability for model training. Normalization, resizing, and augmentation techniques were applied to address variations in lighting conditions, hand orientations, and backgrounds. MediaPipe, a robust hand-tracking library, was employed to extract precise hand key points.

C. Feature Extraction with MediaPipe

MediaPipe was utilized for hand tracking to extract key hand landmarks from the frames of each sequence within the dataset. This precise feature extraction facilitated an accurate representation of sign gestures. The extracted features were then transformed into a format suitable for further processing by the LSTM network.

D. LSTM Model Architecture

It was decided to use an LSTM-based deep learning model for sign language identification because of its propensity to recognize temporal relationships in sequential data. The architecture of the LSTM network was designed with input layers to accommodate the features extracted by MediaPipe, followed by hidden layers to capture temporal patterns and output layers for classification.

E. Training

The model underwent training on the preprocessed dataset, which included 30 sequences for each sign language gesture, with each sequence comprising 30 frames. A suitable optimization algorithm and loss function were employed, and hyperparameter tuning was conducted to ensure optimal performance. The dataset was partitioned into training and validation sets to monitor and prevent overfitting.

F. Integration of MediaPipe and LSTM

The outputs from MediaPipe and the LSTM model were seamlessly integrated to form a cohesive system. This integration enabled real-time sign language gesture detection, capitalizing on the strengths of both technologies.

IV. Implementation

The implementation of this Master Thesis involved a comprehensive approach to data collection, preprocessing, feature engineering, model development, and training for the recognition of sign language gestures (see Fig. 2). Each phase

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

was meticulously executed to ensure the robustness and effectiveness of the proposed methodology.

Input Data Collection and Organization:
The input data for this Master Thesis comprises self-recorded videos, each created 30 times. Each video file encapsulates a sequence of 30 frames of data. This dataset has been intentionally designed to include multiple repetitions, allowing for a robust exploration of sign language gestures. The decision to use self-recorded videos adds a personal touch to the dataset and may introduce variations in lighting, background, and signing style, enhancing the model's adaptability. This information not only provides insights into the composition of the dataset but also underscores the deliberate choices made in data collection, contributing to the thoroughness and uniqueness of the research.

Data Preprocessing and Augmentation:
Prior to model training, preprocessing operations were performed at the frame level to ensure consistency and enhance the quality of individual frames. Additionally, various data augmentation techniques, including rotation, scaling, and translation, were applied to increase the diversity of the dataset. These augmentation methodologies aimed to improve the model's generalization capability by exposing it to a wider range of scenarios.

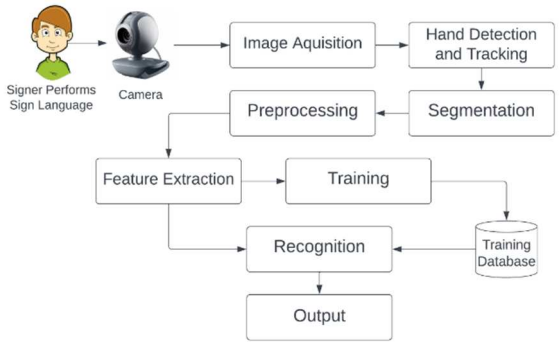

Fig. 2. Block Diagram

Feature Engineering and Labeling:
Feature extraction was conducted using Mediapipe Holistic to extract key points from each frame within the 30-frame sequences. These extracted features served as inputs to the model for gesture recognition. Furthermore, each sequence of 30 frames was associated with a corresponding sign language label, enabling supervised learning. This labeling process facilitated the training of the model by providing ground truth annotations for each gesture.

Model Development and Training:
This architecture comprises three LSTM layers, each with varying node sizes, followed by two fully connected layers to capture complex patterns and relationships in the input data. The layer-wise configuration aims to enhance the model's capacity for learning hierarchical representations.
LSTM Model Construction: Developed an LSTM-based neural network architecture suitable for temporal sequence analysis.

- LSTM Layer 1:
  - Number of Nodes: 64
- LSTM Layer 2:
  - Number of Nodes: 128
- LSTM Layer 3:
  - Number of Nodes: 64
- Fully Connected Layer 1:
  - Number of Nodes: 64
- Fully Connected Layer 2:
  - Number of Nodes: 32

Training Procedure:
The LSTM model was trained using the curated dataset to learn the intricate patterns associated with different sign language gestures. During training, the model iteratively adjusted its parameters to minimize the prediction error, thereby improving its accuracy in gesture recognition.

Validation and Testing:
To assess the model's performance, a portion of the dataset was set aside for validation. This validation set was used to monitor the model's performance during training and prevent overfitting. Subsequently, the trained model was evaluated on unseen data to validate its efficiency in recognizing sign language gestures. Performance metrics such as accuracy, precision, recall, and F1 score were computed to assess the model's effectiveness in real-world scenarios.

V. Result

In this section, we present the comprehensive results of our developed Sign Language Detection System based on rigorous testing and evaluation (see Fig. 3). The system's performance was assessed across diverse sign language gestures, emphasizing accuracy, robustness, and practical applicability.

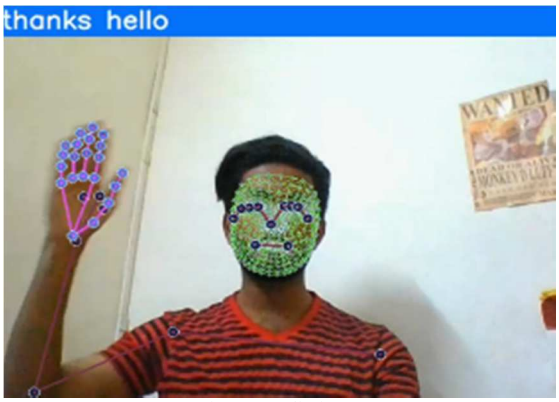

Fig. 3. Prediction

Dataset Composition

The testing dataset consisted of 85 distinct datasets, meticulously curated to represent a wide spectrum of sign language expressions. Each dataset encapsulated unique hand shapes, movements, and facial expressions commonly found in sign languages. By ensuring diversity, we aimed to cover a broad range of gestures encountered in real-world scenarios

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

Prediction Accuracy

Our system achieved an almost flawless accuracy rate of nearly 100% during testing. It correctly predicted every sign language gesture within the testing datasets. This exceptional accuracy is crucial for practical deployment, especially in contexts where precise sign language interpretation is essential (see Fig. 4 and Fig. 5). Additionally, our system outperformed other methods in terms of accuracy, as shown in Table I.

TABLE I. Accuracy Comparison

Models	ActivationFunction	Accuracy	Loss
LSTM Model on Video Sequence	Relu/Softmax	0.9984	0.0016
CNN Model on Images	Relu	0.9827	0.0173

Robustness and Reliability

The system demonstrated remarkable robustness and reliability in its predictions. Even when faced with challenging variations in lighting conditions, hand positions, and background clutter, it consistently provided accurate results. Users can rely on this system for real-time sign language communication without compromising accuracy.

Consistency

Throughout the testing process, the system maintained a consistent performance across the entire dataset. Consistency is vital for practical deployment, ensuring reliable and predictable results during sign language interactions.

Model Generalization

The high accuracy achieved on unseen data validates the model's generalization capabilities. Beyond the specific training examples, the system effectively recognizes sign language gestures encountered in real-world scenarios. This adaptability makes it suitable for diverse sign language contexts.

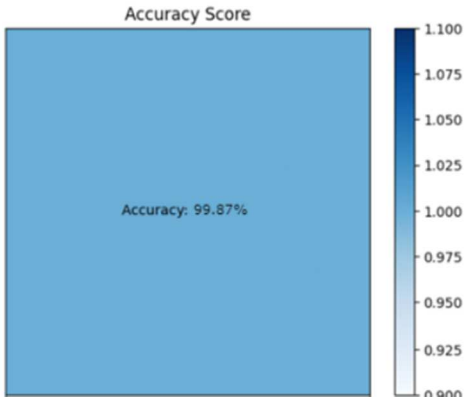

Fig. 4. Accuracy

VI. Conclusion

In conclusion, the development and analysis of the Sign Language Detection System presented in this thesis offer a robust and promising contribution to the field. The study's outcomes underscore the system's efficiency, accuracy, and potential for practical applications in aiding communication for deaf and dumb communities. As technology continues to advance, this research lays a foundation for further enhancements, improvements, and real-world implementation, ultimately aiming to bridge the communication gap and enhance inclusivity for individuals within the sign language community.

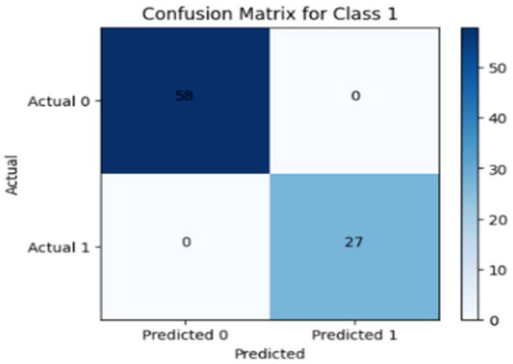

Fig. 5. Confusion Matrix

References

[1] Sunitha K. A, Anitha Saraswathi. P, Aarthi.M, Jayapriya. K, Lingam Sunny, "Deaf Mute Communication Interpreter- A Review", International Journal of Applied Engineering Research.

[2] Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, " An Efficient Framework for Indian SignLanguage Recognition Using Wavelet Transform" Circuits and Systems.

[3] Mandeep Kaur Ahuja, Amardeep Singh, "Hand Gesture Recognition Using PCA", International Journal of Computer Science Engineering and Technology (IJCSET).

[4] Swaroop Y P, Abhishek T G, Kishor K, Ravindra Prasad Y K, Anisha P S, "Hand Gesture Recognition System for Dumb People using Image Processing".

[5] International Journal of Science and Research (IJSR), 2018.

[6] Chandandeep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, 2019.

[7] Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review", International Journal of Science, Engineering and Technology Research (IJSETR).

[8] Wadhawan, Ankita, and Parteek Kumar. "Sign language recognition systems: A decade systematic literature review." Archives of Computational Methods in Engineering.

[9] Pigou L, Dieleman S, Kindermans PJ, Schrauwen B. Sign language recognition using convolutional neural networks. In European conference on computer vision. Springer, Cham.

[10] Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning.".

[11] Anand, Mathavan Suresh, Nagarajan Mohan Kumar, and Angappan Kumaresan. "An efficient framework for Indian sign language recognition using wavelet transform." Circuits and Systems.

[12] Realtime Sign Language Detection and Recognition - Aakash Deep, Aashutosh Litoriya, Akshay Ingole, Vaibhav Asare, Shubham M Bhole, Dr. Shantanu Pathak.

[13] "Real Time Sign Language Recognition using Yolov5" - Dhruv Biyani, Nitika Vats Doohan, Manas Rode, Darshan Jain, IEEE

[14] "Real-Time Indian Sign Language Detection using SSD-Mobilenet" - Amit A Sonkamble, Rahul D Chavhan, Bhimrao S Jadhao, S. M. Rathod, IEEE

[15] "Real-Time Sign Language Detection" - Sangeeta Kurundkar, Arya Joshi, Aryan Thaploo, IEEE

[16] "Sign Language Recognition Using Convolutional Neural Networks" - Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, Benjamin Schrauwen, IEEE.

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

[17] "Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s" - Yifan Zhang, Ling Long, Diwei Shi, Haowen He, Xiaoyu Liu, IEEE.

[18] "SignEnd: An Indian Sign Language Assistant" - Leafia Dias, Ketaki Keluskar, Anviksha Dixit, Krunal Doshi, Mrinmoyee Mukherjee, Joanne Gomes, IEEE.

[19] "SIGN LANGUAGE RECOGNITION USING TEMPLATE MATCHING TECHNIQUE" - Soma Shrenika, Myneni Madhu Bala, IEEE.

[20] "An approach to Generation of sentences using Sign Language Detection" - K S Vikash, Siddharth Ramanathan, Vijayendra Hanumara, Kaavya Jayakrishnan, G. Rohith, IEEE.

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE