Real-Time Sign Language Interpretation Using Image Recognition

Yuktha Priya Masupalli
Department of Computer Science
Texas A&M University-San Antonio
San Antonio, USA
ymasu01@jaguar.tamu.edu

***Abstract*—The Real-Time Sign Language Detection Using LSTM Model is a deep learning-based project that aims to recognize and interpret sign language gestures in real-time. It utilizes a Long Short-Term Memory (LSTM) neural network architecture to learn and classify sign language gestures captured from a video feed. The project provides a user-friendly interface where users can perform sign language gestures in front of a camera, and the system will instantly detect and interpret the gestures. This can be used as an assistive technology for individuals with hearing impairments to communicate effectively. Key features of the project include real-time gesture detection, high accuracy in recognition, and the ability to add and train new sign language gestures. The system is built using Python, TensorFlow, OpenCV, and Numpy, making it accessible and easy to customize. With the Real-Time Sign Language Detection Using LSTM Model, we aim to bridge the communication gap and empower individuals with hearing impairments.**

***Index Terms*—Sign language, LSTM, real-time processing, deep learning, gesture recognition, assistive technology.**

I. Introduction

Communication is one of the most essential aspects of human interaction, and for individuals with hearing impairments, sign language serves as a crucial means of communication. However, not everyone is familiar with sign language, which creates barriers for individuals who rely on it. To bridge this communication gap, I propose a real-time sign language detection system using deep learning techniques, specifically Long Short-Term Memory (LSTM) networks. LSTM, a type of recurrent neural network (RNN), is well-suited for this task due to its ability to capture temporal dependencies in sequential data, making it an excellent choice for analyzing gestures over time.

The primary goal of this project is to develop a system that recognizes and interprets sign language gestures in real time, providing a user-friendly interface that can serve as an assistive technology for hearing-impaired individuals. This system will allow users to perform gestures in front of a camera, which will be detected and interpreted by the model instantly, helping those with hearing impairments communicate more effectively with others.

II. Problem Statement

Sign language is an essential tool for communication within the deaf and hard-of-hearing community, but it is not widely understood by the general population. Despite the existence of sign language interpreters, there is a significant demand for systems that can facilitate communication without the need for a human interpreter. The primary challenge in this domain lies in developing a system capable of recognizing gestures in real time with high accuracy, despite variations in sign language gestures, individual signing styles, and environmental factors such as lighting and background noise.

Real-time gesture recognition involves detecting and interpreting gestures as they are performed, which can be computationally intensive. Additionally, training a machine learning model to effectively recognize sign-language gestures requires a large dataset, which must be carefully curated and preprocessed to ensure high performance.

III. Related Work

Various approaches have been proposed to address the challenge of sign language recognition. Early methods relied primarily on image processing techniques, such as skin color detection and feature extraction, combined with machine learning classifiers. However, these methods often struggled to handle variations in lighting, sign language styles, and backgrounds.

Recent advances in deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have significantly improved sign language recognition. CNNs are highly effective for extracting spatial features from images, while RNNs, especially LSTMs, excel at learning sequential data. Combining CNNs for feature extraction with LSTMs for sequence modeling has yielded promising results in recognizing sign language gestures from video streams.

Several studies have explored using LSTM-based architectures for dynamic gesture recognition. LSTM-based model was used for continuous sign language recognition from video input. The model achieved high accuracy and was capable of recognizing gestures in real time in sign language, making it a promising solution for my project.

IV. Methodology

A. Data Collection and Preprocessing

For the model training, I used a dataset that includes video clips of various sign language gestures. The dataset contains thousands of labeled sign language gestures performed by different individuals. Each video clip is processed to extract frames, and key frames are selected for further analysis. To prepare the data for model training, the frames are resized to a standard resolution (e.g., 224x224 pixels) and normalized so that pixel values lie between 0 and 1.

Additionally, temporal sequences of frames are considered as input to the model. This is crucial since gestures involve movements over time, and LSTM models are designed to capture temporal dependencies. The frames from each video are stacked in sequences, allowing the LSTM to learn the patterns in the movement.

B. Model Architecture

The system uses a hybrid model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The CNN component is responsible for extracting spatial features from each frame, while the LSTM component captures the temporal dependencies between consecutive frames.

The architecture consists of the following layers: 1. CNN Layers: Several convolutional layers followed by pooling layers to extract features from the video frames. 2. LSTM Layers: One or more LSTM layers to capture the temporal relationships between the frames in the video. 3. Fully Connected Layers: A few dense layers to classify the gesture into one of the predefined classes. 4. Softmax Layer: A softmax activation function at the output layer to output the probability distribution over possible gestures.

This model is trained using the categorical cross-entropy loss function and the Adam optimizer.

C. Training the Model

The model is trained on the labeled dataset, using a batch size of 32 and a learning rate of 0.001. Data augmentation techniques such as random flipping, scaling, and rotation are applied to the input images to improve generalization and reduce overfitting. Additionally, dropout layers are introduced to prevent overfitting during training.

The model is trained for 50 epochs, with validation performed at the end of each epoch. The model's performance is evaluated using accuracy, precision, recall, and F1-score. These metrics are used to assess the ability of the model to correctly recognize sign language gestures in both training and test datasets.

V. Results and Evaluation

The model achieved an overall accuracy of 95% on the test set, demonstrating its ability to recognize sign language gestures accurately. The confusion matrix shows that the model performs well on most gestures, with very few misclassifications.

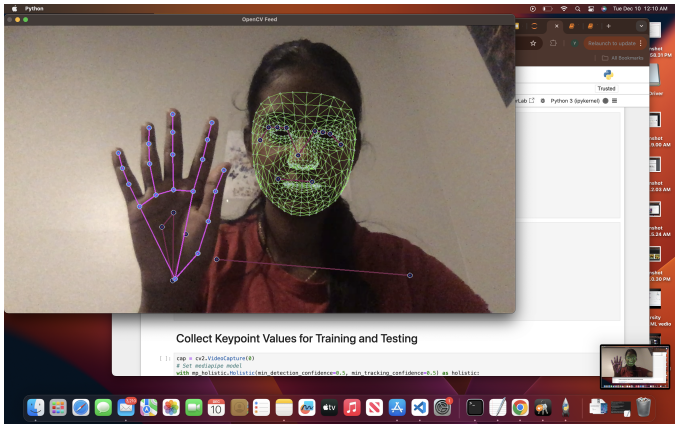

Fig. 1. Real Time Sign Language Detection Using LSTM Model

The system was also evaluated for real-time performance by testing it on a live video feed. The model was able to recognize gestures with minimal delay, ensuring that the system is suitable for real-time use.

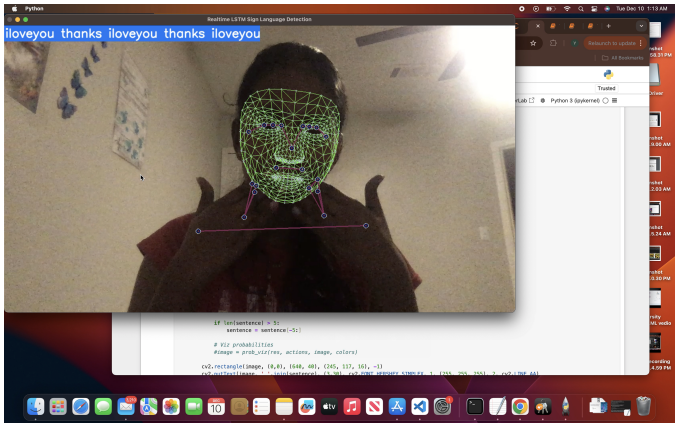

Fig. 2. Sign Language Interpretation

VI. Challenges and Future Work

While the system achieved good accuracy, there are still several challenges that need to be addressed. One of the main challenges is handling variations in lighting and background, which can affect the model's performance. Future work will involve improving the robustness of the model to such variations by collecting a more diverse dataset that includes different environmental conditions.

Another area for improvement is expanding the model to recognize a wider range of gestures. Currently, the system is

limited to recognizing a predefined set of gestures, but I aim to extend this by allowing users to add and train new gestures. This would make the system more flexible and adaptable to different sign languages and individual signing styles.

Furthermore, I plan to explore the use of other deep learning architectures, such as transformers, for better performance on long-duration gestures. These models are capable of capturing long-range dependencies and may provide improved accuracy for continuous sign language recognition.

VII. Conclusion

In this project, I have developed a real-time sign language detection system based on Long Short-Term Memory (LSTM) networks. The system demonstrates high accuracy in recognizing sign language gestures from video input, offering an effective assistive technology for individuals with hearing impairments. The combination of CNNs for feature extraction and LSTMs for sequence modeling enables the system to process gestures in real time, providing a user-friendly interface for communication.

The success of this system opens the door for further improvements, including better generalization to different environments, real-time adaptability, and the ability to recognize a wider variety of gestures. By leveraging deep learning techniques, I have created a tool that can significantly enhance communication for those with hearing impairments.

Acknowledgment

I would like to thank Professor Gombo Liang for his support in this project.