Sign Language Recognition App with Python

Nandini Singh
Department of Artificial Intelligence and Data Science,
Dr. Akhilesh Das Gupta Institute of Technology and Management, New Delhi
Email: nandinisingh5may@gmail.com

***Abstract* - This project was developed with the aim of utilizing computer vision techniques to achieve real-time and highly accurate Sign Language recognition (SLR). Its purpose is to bridge the communication gap between individuals with varying levels of hearing abilities. To accomplish this, a dataset comprising images corresponding to signs was created and labeled accordingly. These images were then processed through a neural network that leverages transfer learning. Moreover, the project incorporates Mediapipe Holistic and the system utilizes Long Short-Term Memory (LSTM) architecture to model the temporal dynamics of the sign language gestures. Through the integration of these technologies, the project endeavors to enable accurate and real-time recognition of sign language.**

Keywords - **Sign Language Recognition, Deep Learning, Computer Vision, Image Preprocessing, Transfer Learning, LSTM Neural Network, MediaPipe Holistic.**

INTRODUCTION

Sign language serves as a crucial mode of communication for individuals to express their thoughts and emotions non-verbally. However, there is a significant communication barrier between those familiar with sign language and those who are not, leading to the marginalization of individuals with hearing impairments. Sign language heavily relies on hand movements and facial expressions to convey meaning. While the hearing-impaired community frequently utilizes sign language for communication, it is rarely understood by those without hearing impairments. Consequently, social interactions are severely limited, and relying on real-time interpreters can be both impractical and costly. In this project, we will create a program using OpenCV to recognize and translate custom sign language gestures into regular text, aiming to bridge the communication gap and enable better understanding and inclusivity.

Fig 1: Skin Masked Images of different English Alphabets

1.1 Applications

A Python-based sign language recognition application has diverse practical uses in various fields such as communication, education, accessibility, and interactive experiences. One of its primary purposes is to enhance communication for individuals who are deaf or hard of hearing by converting sign language gestures into text or speech, enabling effective interaction with others. Moreover, the app serves as a valuable tool for learning sign language, providing visual feedback and explanations to aid in the learning process. By incorporating augmented reality capabilities, the application utilizes camera input to recognize gestures and overlay virtual elements, enhancing the user experience. Lastly, the app can be utilized for interactive sign language-based games and applications, promoting engagement and interactive learning. Overall, this application actively promotes inclusivity, accessibility, and effective communication for individuals with diverse communication needs.

1.2 Role of different fields

The technology and software development industry play a vital role in the creation of the sign language recognition app by providing the necessary tools and frameworks for app development. Ongoing research and development efforts in fields like computer vision, machine learning, and natural language processing contribute to enhancing the app's algorithms and improving gesture recognition capabilities. Collaboration with accessibility and inclusion organizations ensures that the app is designed to meet the specific needs of individuals with hearing impairments, incorporating their valuable guidance and support. Educational institutions and language learning platforms also play a crucial role by partnering to develop comprehensive sign language databases and effective learning methodologies, enhancing the app's functionality and educational value. Through these collective efforts, the technology, software development, accessibility, and education sectors collaborate to create a sign language recognition app that promotes accessibility, inclusion, and effective communication.

1.3 Recent Advancements in Sign Language Detection

In recent years, sign language detection has witnessed notable advancements driven by the application of convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These deep learning models excel in capturing spatial and temporal patterns inherent in sign language gestures, leading to enhanced accuracy and robustness. The availability of extensive sign language datasets has played a crucial role in training these models, enabling them to learn intricate features and accommodate variations within sign language gestures. Furthermore, the incorporation of computer vision techniques, such as optical flow analysis and hand tracking algorithms, has contributed to the refinement of sign language detection systems, offering more precise and reliable tracking of hand movements. These combined developments in deep learning, increased data availability, and computer vision techniques have significantly elevated the performance of sign language detection and recognition systems, opening doors to improved communication and accessibility for individuals who use sign language.

1.4 Challenges

Developing a sign language recognition app comes with its own set of challenges. One major hurdle is the inherent variability present in sign language gestures. Sign language encompasses a wide range of movements and gestures that can vary across regions, individuals, and contexts. Accurately capturing and recognizing this variability requires robust algorithms and extensive training data to encompass the diverse range of possible gestures. The dynamic nature of sign language, including hand movements, facial expressions, and body language, adds another layer of complexity to the recognition process. Ensuring real-time and accurate detection of these dynamic features presents a significant challenge. Additionally, environmental factors such as lighting conditions and background clutter can impact the app's performance by interfering with gesture detection. Overcoming these challenges necessitates ongoing research, algorithm refinement, and the collection of diverse and representative datasets to train the models effectively.

2. LITERATURE REVIEW

Several research papers have contributed to advancements in sign language recognition. In recent research [1], a deep learning-based pipeline was developed to automatically recognize sign language from RGB videos. The pipeline utilized SSD, 2DCNN, 3DCNN, and LSTM, and introduced a novel representation of hand skeleton features. Additionally, 3DCNNs were applied to pixel level and heat map features to enhance discriminative capabilities. In a separate investigation [2], a deep CNN architecture was introduced for detecting and classifying sign languages, incorporating both static and dynamic gestures in the training process. Furthermore, a study [3] conducted a comparative analysis of machine learning and deep learning models for classifying American sign language, ensuring robustness through user-independent k-fold cross-validation and test phases. Another study [4] focused on real-time sign language recognition and proposed a model that combined a single shot detector, 2D CNN, SVD, and LSTM. Another approach [5] addressed human action recognition, utilizing motion tracking and feature extraction through a Recurrent Neural Network model with Gated Recurrent Unit. These advancements hold great promise for improving communication, accessibility, and inclusion

3. METHODOLOGY

This research proposes a method for sign language detection using OpenCV and machine learning techniques. The proposed method consists of the following steps:

3.1 General Design

In our sign language recognition project, we have developed a sign detector capable of detecting custom signs and easily expanding to include a wide range of additional signs and hand motions, such as the alphabet and numerals. To build this project, we utilized Python modules including OpenCV, Mediapipe, Tensorflow, and Keras. The OpenCV module processes live video frames from a camera, analyzing the actions of the person being displayed. Mediapipe Holistic is used to extract keypoints from the hands, torso, and face in the video frames. These keypoints are then fed into the prediction algorithm, which begins the real-time prediction of the sign being made. The predicted sign is displayed as the expected output.

3.2 Prerequisites

The prerequisites software & libraries:

1. Python (3.10.8)
2. IDE (Jupyter)
3. Mediapipe (version 0.10.1)
4. Numpy (version 1.23.5)
5. cv2 (openCV) (version 4.7.0.72)
6. Keras (version 2.12.0)
7. Tensorflow (version 2.12.0)

3.3 Dataset

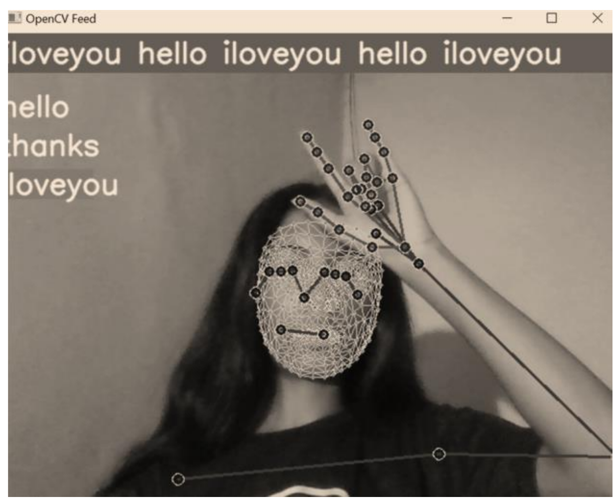
The dataset includes different motions and signals for sign language. A live video stream from the camera is constantly monitored, and frames within the defined region of interest (ROI) that show gestures or motions are saved in a dedicated directory called the gesture directory. Each sign is captured through approximately 30 video sequences, where each sequence contains 30 frames capturing significant moments. These frames are stored as Numpy arrays. The main goal is to accurately identify the specific gesture being performed throughout the entire video sequence.

3.4 Training

In Python, the training process utilizes the TensorFlow machine learning platform. To apply transfer learning, the datasets and label files are preprocessed into a format compatible with TensorFlow. Specifically, tfrecord files are generated from the training and testing data folders. Prior to training, the models designed for object detection need to be downloaded and adjusted in the configuration files to match the number of classes in the dataset. Achieving high accuracy typically requires around 2000 training steps. By combining TensorFlow and Keras, an LSTM model is developed to predict actions displayed on the screen, such as Sign Language gestures in this particular example.

3.5 *Testing*

To evaluate the performance of our model, extensive testing was conducted using deep neural networks and the Mediapipe Holistic framework. The models were trained to anticipate signals based on forearm, hand, and finger kinematics. Among the various configurations tested, the Mediapipe LSTM model with data augmentation demonstrated the highest accuracy, achieving an impressive 100 percent on the test sets. With this successful outcome, our sign language detector is capable of accurately recognizing and detecting hand movements, producing corresponding coordinates, and comprehending sign language gestures. The system operates in real-time, providing immediate and up-to-date signage interpretation.



4. CONCLUSION

In conclusion, the research aimed to create a real-time sign language detection and translation system using deep neural networks and Mediapipe Holistic. The obtained results were promising, indicating the potential for practical applications. The developed system achieved a high accuracy rate, with the Mediapipe LSTM model combined with data augmentation performing exceptionally well, reaching 100% accuracy on test sets. This system effectively recognizes and interprets hand gestures, providing coordinated outputs in real time.

5. FUTURE SCOPE

For future directions, the project team proposes the possibility of developing a mobile application that can classify complete word symbols by incorporating facial emotions and relative hand movements from the face. Furthermore, there are plans to expand the sign language dataset by incorporating common words in addition to alphabet letters. The aim is also to enhance the model's accuracy by introducing additional hyperparameters. These future steps will contribute to further advancements in sign language recognition and translation technology.

6. REFERENCE

[1] Razieh Rastgooa, Kourosh Kiania, Sergio Escalerab(2020), “Hand sign language recognition using multi-view hand skeleton”, Electrical and Computer Engineering Department, Semnan University, Department6of Mathematics and Informatics, Universitat de Barcelona and Computer Vision Center, Barcelona, Spain, , pp.16-24.

[2] R. Bhadra and S. Kar(2021), "Sign Language Detection from Hand Gesture Images using Deep Multi-layered Convolution Neural Network," 2021 IEEE Second International Conference on Control, Measurement and Instrumentation, pp.196-200.

[3] Neziha Jaouedia, Noureddine Boujnahb, Med Salim Bouhlelc(2020), “A new hybrid deep learning model for human action recognition.”, National Engineering School, Avenue Omar Ibn El Khattab Zrig Eddakhlania, Gabes 6072, pp. 447-453.

[4] Rastgoo, R., Kiani, K. & Escalera(2020), “Real-time isolated hand sign language recognition using deep networks and SVD.” J Ambient Intell Human Comput 13, pp. 591–611.

[5] Hernandez V, Suzuki T,(2020), “Convolutional and recurrent neural network for human activity recognition: Application on American sign language” Venture G, pp. 19-23.