Indian Sign Language Detection

Rohith Kumar C N | Rajakumar K | SCOPE

Introduction

Sign languages are systems of visual communication, used primarily by people from deaf communities around the world. In its most common form, words and phrases are signed by gesturing with fingers, hands, arms and facial expressions. Sign languages are fully developed languages with their own grammar and lexicon. Further, they differ from region to region and are often not mutually intelligible with each other, though some languages do possess similarities. They also differ from the spoken languages of a given area in terms of lexicon and rate of articulation. Indian Sign Language (ISL) is a complete language with its own grammar, syntax, vocabulary and other linguistic attributes. ISL varies considerably from its western counterparts in a variety of ways, most noticeably its lexicon, which displays a high level of iconicity. While most other sign languages have a few compound signs (signs consisting of two or more signs), in ISL the compounding system is pervasive.

Motivation

Despite the wide use of sign languages, a communication gap exists between the deaf and mute community and those who do not understand sign languages. This gap often leads to isolation, discrimination, and reduced opportunities for those with hearing and speech impairments. This project aims to bridge this gap by developing an Indian Sign Language Detection system leveraging the power of machine learning and pose detection.

Scope of the Project

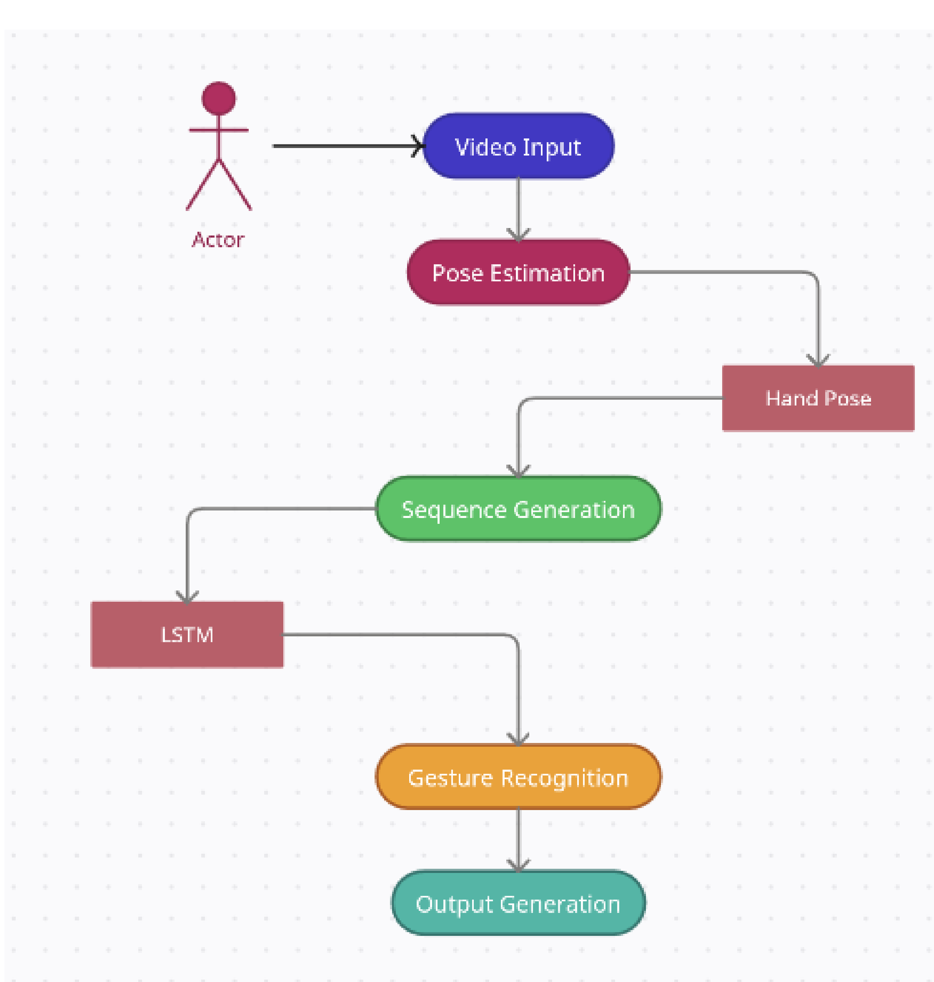
The project will use the MediaPipe Pose Detection model to identify and interpret ISL gestures in real-time, converting them into written text. The translation of these gestures into text allows individuals unfamiliar with ISL to understand and communicate effectively with the ISL user, thus fostering better inclusivity.
By leveraging the power of machine learning and pose detection technology, we can build a system that can translate ISL into text, breaking down barriers and promoting inclusivity. The outcome of this project could revolutionise the way we interact with the deaf and mute community, paving the way for a future where everyone can communicate without barriers.

Methodology

Data Collection:The data collection stage forms the foundation of this project. The performance and generalizability of our machine learning model heavily depend on the quality and variety of the data it is trained on. Our objective is to create a comprehensive dataset of Indian Sign Language (ISL) gestures, which is diverse and representative of real-world scenarios.

Data Preprocessing:Following the data collection, the raw video data must be transformed into a suitable format that can be effectively utilized by our machine learning model. This step is known as data preprocessing and is critical since it can significantly impact the model's performance.By processing each frame of our video data, we will extract a sequence of pose landmarks using mediapose for each ISL gesture. This sequence not only captures the static positioning of the hands but also encapsulates the dynamic motion involved in each gesture. This sequence of pose landmarks will serve as the input data for our machine learning model.

Machine Learning Model:We will use a type of Recurrent Neural Network (RNN) known as Long Short-Term Memory (LSTM) for this task. LSTMs are particularly well-suited for learning from sequential data and have the capability to remember past information, which is crucial in recognizing gestures.The pose landmark sequences will be input into the LSTM model, while the corresponding ISL gestures will be the output. Throughout the training process, the LSTM will learn to recognize patterns in the pose landmark sequences that correspond to each gesture.



Results

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 30, 128)	109056
activation_3 (Activation)	(None, 30, 128)	0
dropout_3 (Dropout)	(None, 30, 128)	0
lstm_3 (LSTM)	(None, 256)	394240
activation_4 (Activation)	(None, 256)	0
dropout_4 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65792
activation_5 (Activation)	(None, 256)	0
dropout_5 (Dropout)	(None, 256)	0
batch_normalization_1 (BatchNormalization)	(None, 256)	1024
dense_3 (Dense)	(None, 262)	67334

Total params: 637,446
Trainable params: 636,934
Non-trainable params: 512

```
Epoch 195/200
85/85 [==============================] - 13s 156ms/step - loss: 0.0780 - accuracy: 0.9800 - val_loss: 1.3005 - val_accurac
y: 0.8163
Epoch 196/200
85/85 [==============================] - 13s 157ms/step - loss: 0.0923 - accuracy: 0.9763 - val_loss: 1.1764 - val_accurac
y: 0.8296
Epoch 197/200
85/85 [==============================] - 14s 160ms/step - loss: 0.1062 - accuracy: 0.9755 - val_loss: 1.3854 - val_accurac
y: 0.8178
Epoch 198/200
85/85 [==============================] - 14s 165ms/step - loss: 0.0767 - accuracy: 0.9815 - val_loss: 1.2401 - val_accurac
y: 0.8207
Epoch 199/200
85/85 [==============================] - 14s 159ms/step - loss: 0.0727 - accuracy: 0.9811 - val_loss: 1.1831 - val_accurac
y: 0.8222
Epoch 200/200
85/85 [==============================] - 13s 154ms/step - loss: 0.0709 - accuracy: 0.9804 - val_loss: 1.1977 - val_accurac
y: 0.8148
2/2 [==============================] - 0s 16ms/step - loss: 0.6013 - accuracy: 0.9143
```

Conclusion

Brief summary of what you discovered based on results

Indicate and explain whether or not the data supports your hypothesis

Give directions for future work in this area.

Each of the areas can be expanded as per requirement. Please use only a two columns to make it visually appealing.

References

The body text / font size should be between 24 and 32 points. Calibri font