Sign Language Recognition

Machine Learning Model to interpret words form human actions.

Problem Statement

In this sign language recognition project, We build a ML model which will take video as input and gives the word which the person is showing. It will be a useful model for the deaf people to convey their thoughts to other people.

Phases of model building

1. Data Collection
2. Feature extraction
3. Model architecture
4. Model training
5. Model testing
6. Real time usage

DATA SET :

- We are using WLASL dataset for the model

- WLASL is a larege video dataset for Word-Level American Sign Language (ASL) recognition, which features 2,000 common different words in ASL.

  Link : https://dxli94.github.io/WLASL/

FEATURE EXTRACTION :

- The input for this is a video. A video is combination of frames
- As the videos are short (<5 sec) we will be breaking the video into frames
- Then extract features from each frame

- To extract features from each frame we use mediapipe framework
- Then we use the series of frames to detect the action
- To extract features from each frame we use mediapipe framework

MEDIAPIPE FRAMEWORK :

It is an open source library to identify the landmarks on hands, face and body pose

1) **Hand:**

LEFT HAND

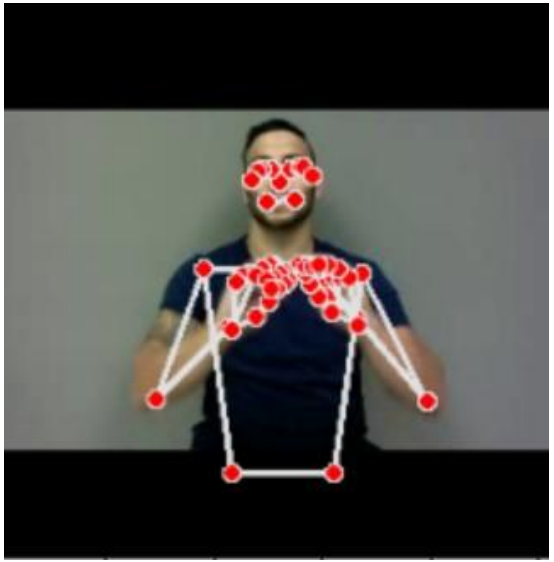
RIGHT HAND

21 POINTS FOR EACH HAND

2) **Pose + Face :**

132 LANDMARKS

Total :



INPUT FORMAT

- Each input video is divided into 30 frames
- Each frame has 258 landmarks (features from hands, pose, face)
- So each input is of form (30*258) 2d vector

DATA LABELING

- We are considering only 10 gestures for this project from the dataset provided
- Actions = { 'book', 'drink', 'computer', 'before', 'chair', 'go', 'clothes', 'who', 'candy', 'cousin' }
- The video will be classified into one of the above actions

MODEL ARCHITECTURE :

- As we are taking a series of frames as input and the frames are interdependent on each other we will use RNN’s

- We will use LSTM for making the model

LSTM Model

- **Long short-term memory** (**LSTM**) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning **(DL)**. Unlike standard feedforward neural networks, LSTM has feedback connections

- LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs

Model: "sequential"

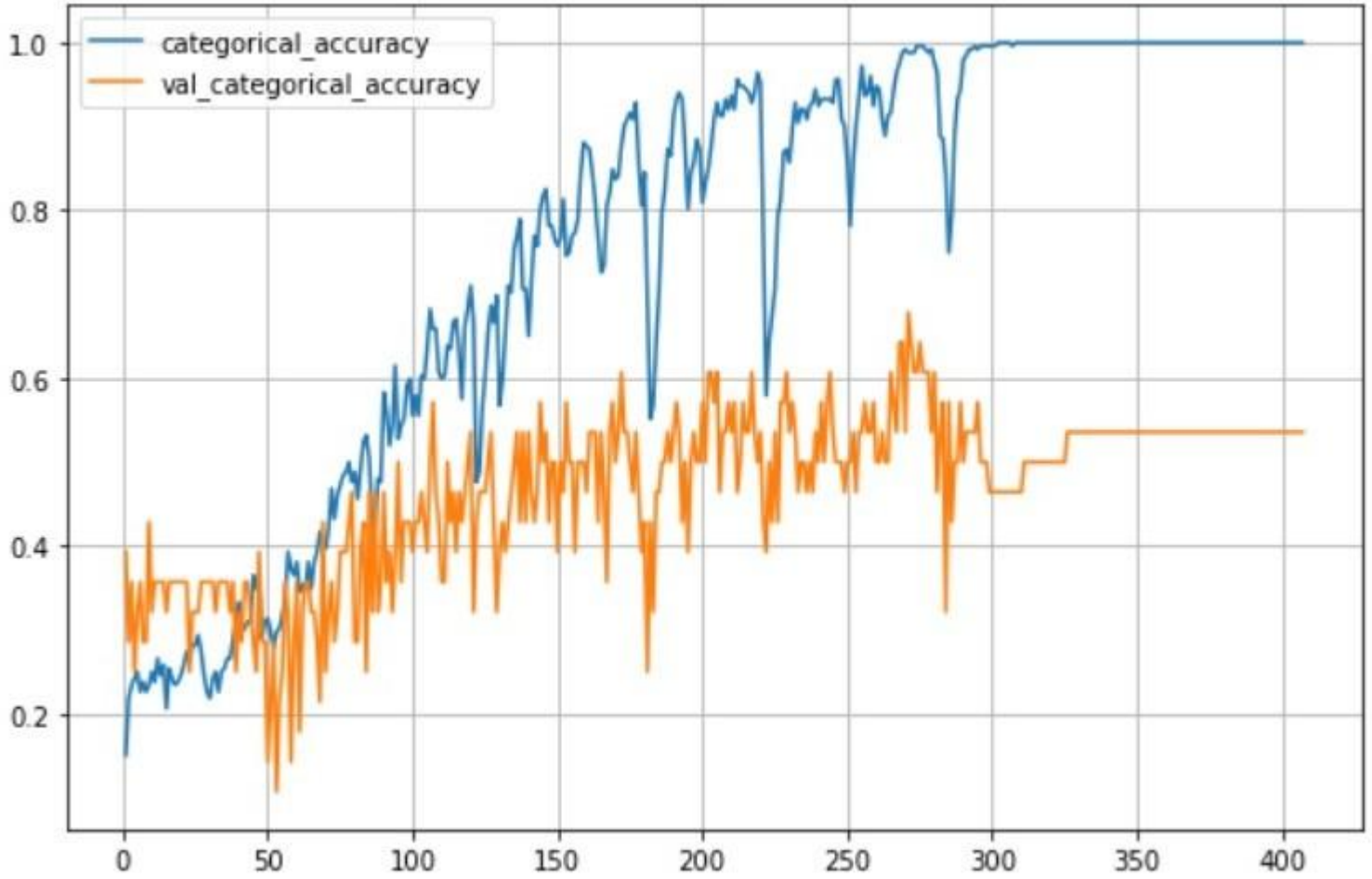
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 32)	37248
lstm_1 (LSTM)	(None, 30, 128)	82432
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 10)	330

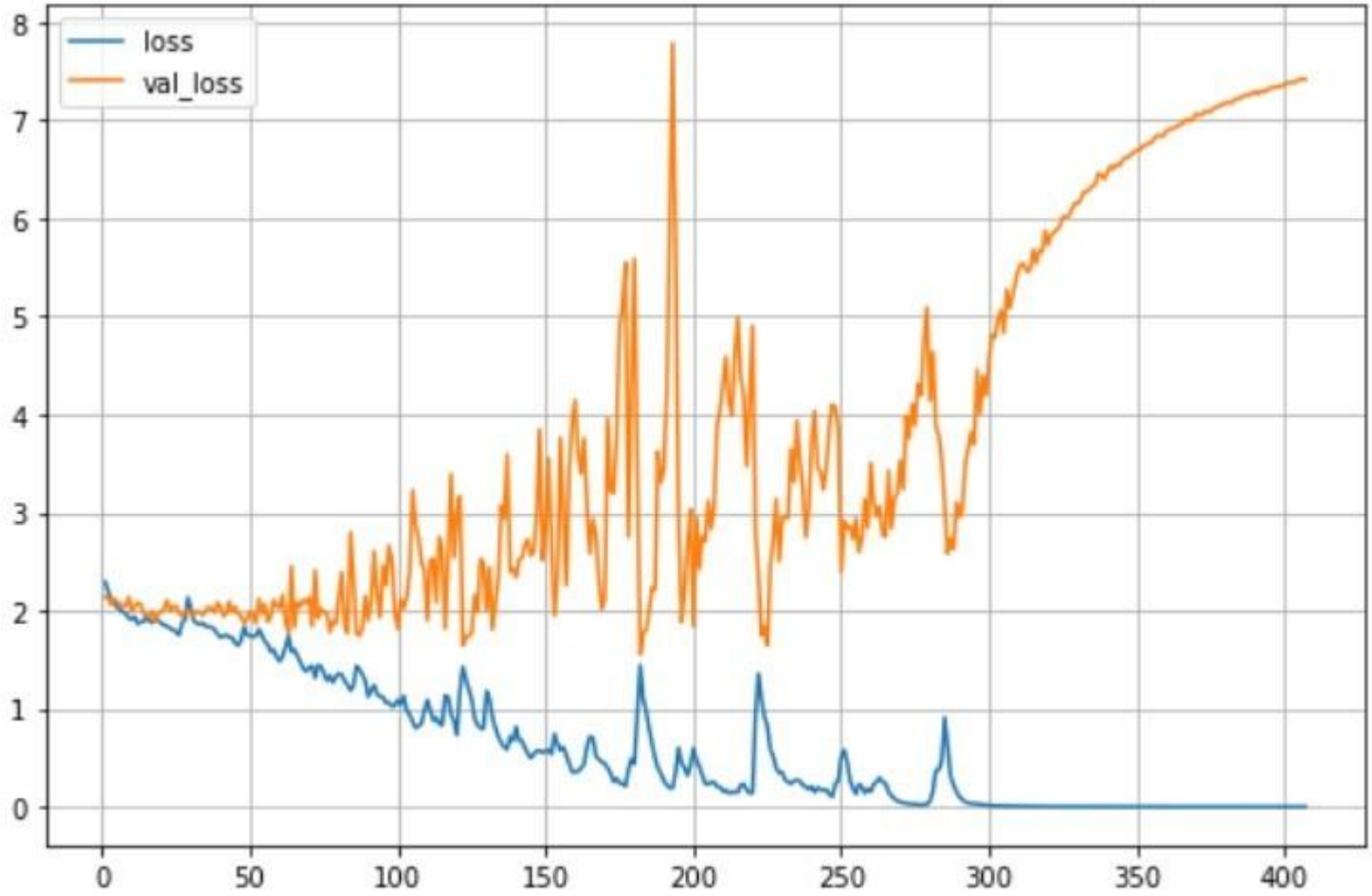
Total params: 175,658
Trainable params: 175,658
Non-trainable params: 0

ACCURACY

Adam:

0.5357142857142857


categorical_accuracy
val_categorical_accuracy
1.0
0.8
0.6
0.4
0.2
0
50
100
150
200
250
300
350
400


loss
val_loss
8
7
6
5
4
3
2
1
0
0
50
100
150
200
250
300
350
400