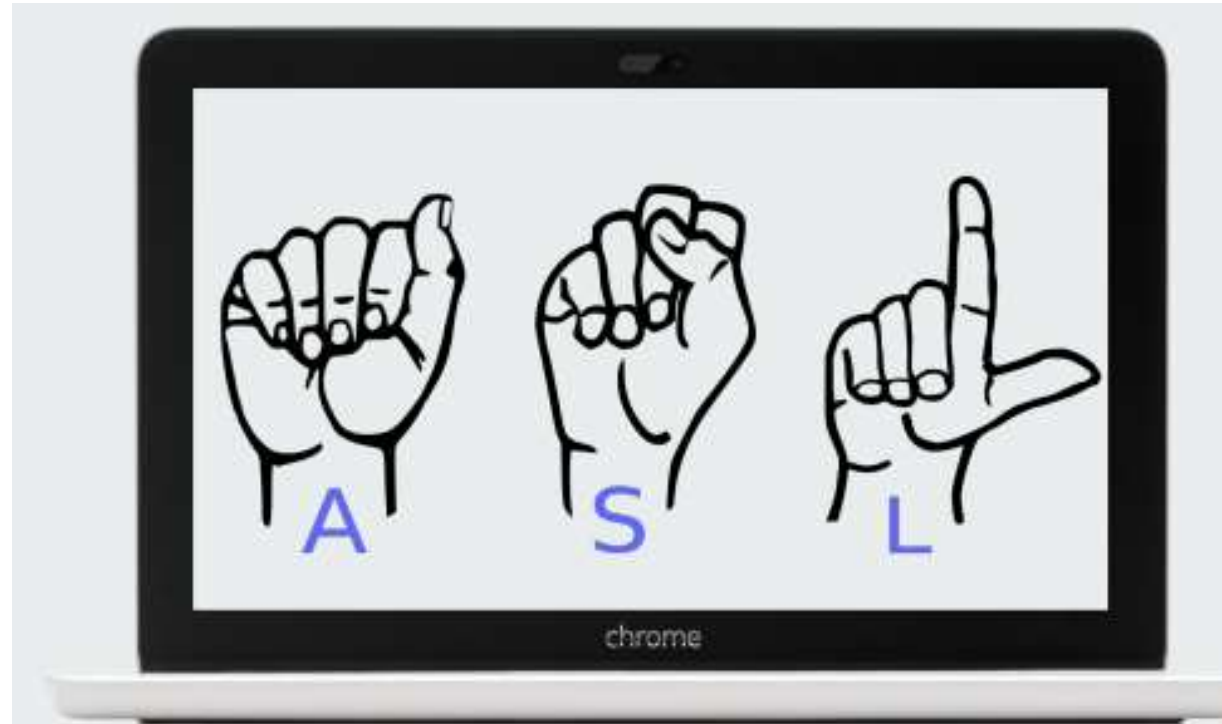# Conversion of Sign Language to Text

For Deaf And Dumb

Abstract

▶ Our project aims to create a computer application and train a model which when shown a real time video of hand gestures of American Sign Language shows the output for that particular sign in text format on the screen.

Sign language is Visual language and consists of 3 major components:

Fingerspelling	Word level Vocabulary	Non-manual features
Used to spell word Letter by letter	Used for majority of conversation.	Facial Expression and tongue, mouth and body position

We implemented 27 symbols(A-Z) of ASL in our project.
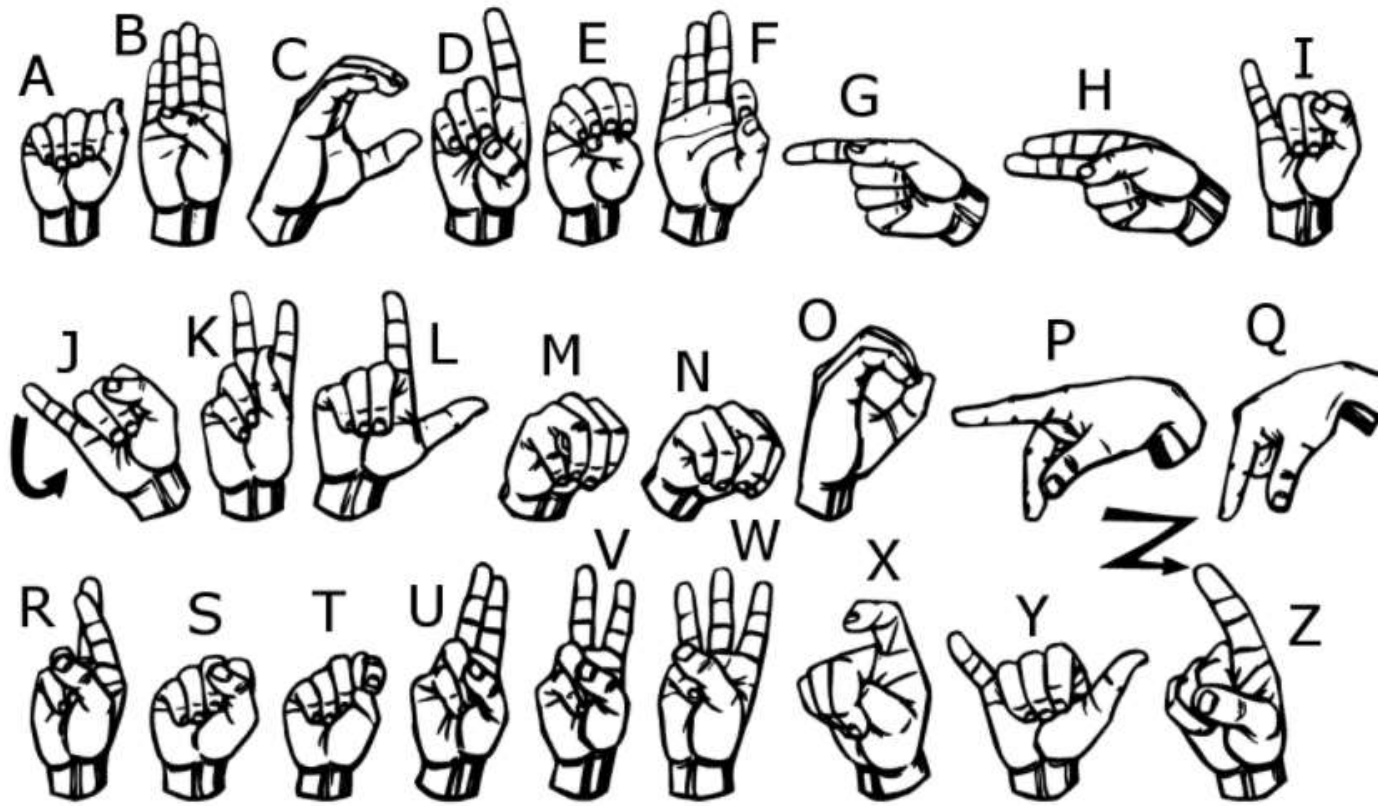
How did we generated Data set and did Data Processing ?

- ▶ We Created our own Data Set because we couldn't find dataset in the form of raw Image that matched our own requirement .

- ▶ All we could find were the Dataset in the RGB values.

- ▶ Hence we Decided to Created our own Data set.

- ▶ The Dataset in made using OpenCV Image Processing by converting the image in the Gray scale and then Gaussian Blur.

- ▶ The capture Image is then labelled and Stored in the train and test as Dataset Files.

Convolutional Neural Network

▶ CNNs consist of multiple convolutional layers each layer containing numerous "filters" which perform feature extraction.

▶ Initially these "filters" are random and by training, the feature extraction gets better by better.

▶ It's primarily used for image classification.

▶ After Convolutional Layer is the Max Pooling Layer Which Reduce the Image Size of the Convolutional Layer .

▶ There the Flatten Layer which convert the multidimensional array to Single dimension to be passes to the dense layer.

▶ The Dense layer consists of Ann to be processed .

▶ The Last layer consist of Output Dense layer which has output Prediction.
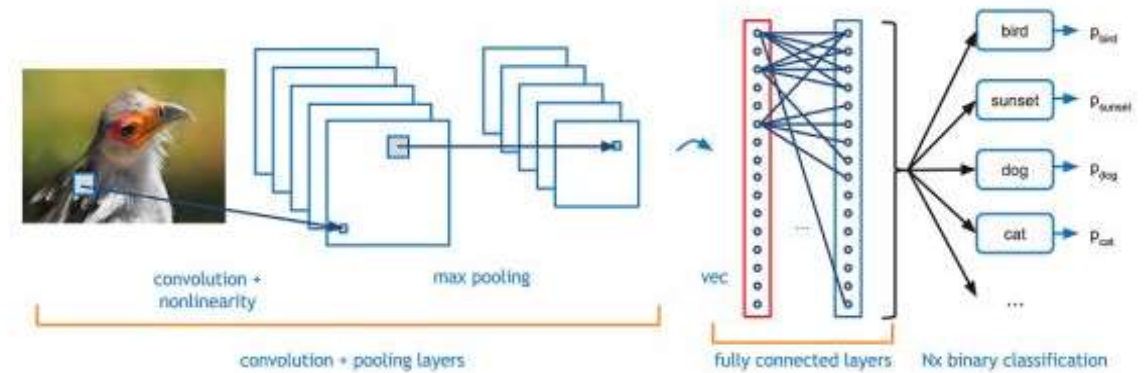

Figure 5.2: Convolution neural networks

Gesture Classification

▶ Gesture Classification is Done By layer One and The First Prediction is Done if the prediction is Have Similarities which Done Gesture then output is then Send to next layer two to get the precision output of the gesture.

▶ Algorithm Layer 1:

▶ 1. Apply gaussian blur filter and threshold to the frame taken with OpenCV to get the processed image after feature extraction.

▶ 2. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word.

▶ Algorithm Layer 2:

▶ 1.We detect various sets of symbols which show similar results on getting detected.

▶ 2. We then classify between those sets using classifiers made for those sets only.

▶ In our testing we found that following symbols were not showing properly and were giving other symbols also :

▶ 1. For D : R and U

▶ 2. For U : D and R

▶ 3. For I : T, D, K and I

▶ 4. For S : M and N

Challenges Faced

▶ We couldn't find a dataset with raw images of all the asl characters so we made our own dataset.

▶ Second issue was to select a filter for feature extraction. We tried various filter including binary threshold, canny edge detection, gaussian blur etc. ,of which gaussian blur filter was giving better results.

▶ Issues were faced relating to the accuracy of the model we trained in earlier phases which we eventually improved by increasing the input image size and also by improving the dataset.

Software Requirements

- Python 3.10.4

- TensorFlow 2.9.0

- Matplotlib 3.5.2

- OpenCV 4.6.0.66

- Keras 2.9.0

- Jupyter 1.0.0

- Notebook 6.4.12

Limitations of our model

- The model works well only in good lighting conditions.
- Plain background is needed for the model to detect with accuracy.

Conclusion and Future Scope

▶ In this report, a functional real time vision based American sign language recognition for D&M people have been developed for asl alphabets.

▶ We achieved an accuracy of 99.57% on our dataset.

▶ Prediction has been improved after implementing two layers of algorithms in which we verify and predict symbols which are more similar to each other.

▶ We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms.

▶ We are also thinking of improving the preprocessing to predict gestures in low light conditions with a higher accuracy.

Project Under the Supervision of Prof. Javed Pathan

Effort by

Deepak Gond

Thank You……….!