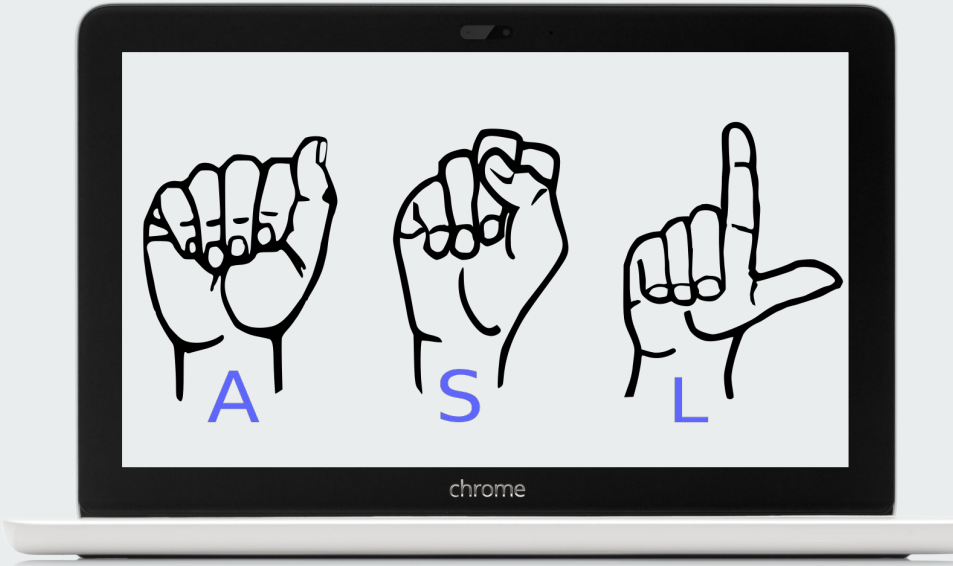
Hand Gesture Recognition using CNN

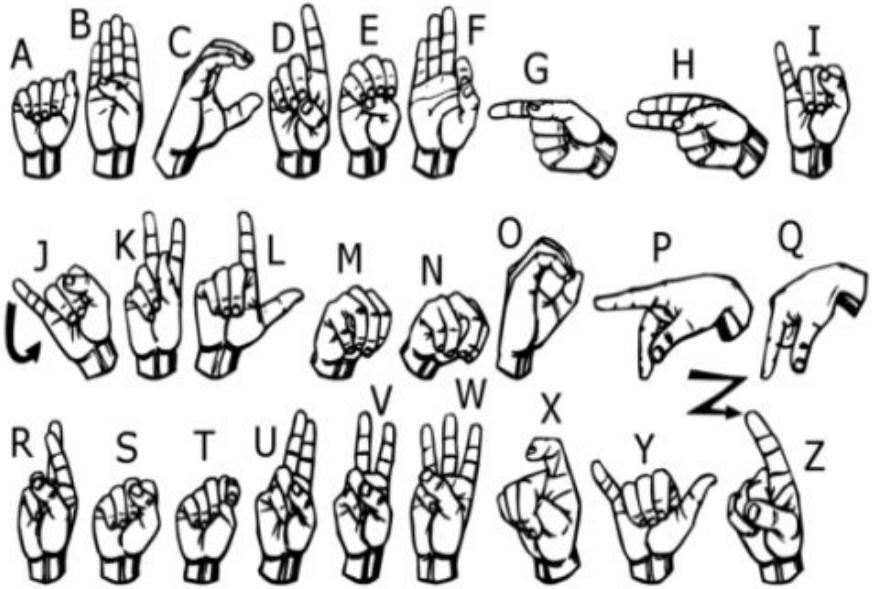


ASL - American Sign Language
CNN - Convolutional Neural Network

Abstract

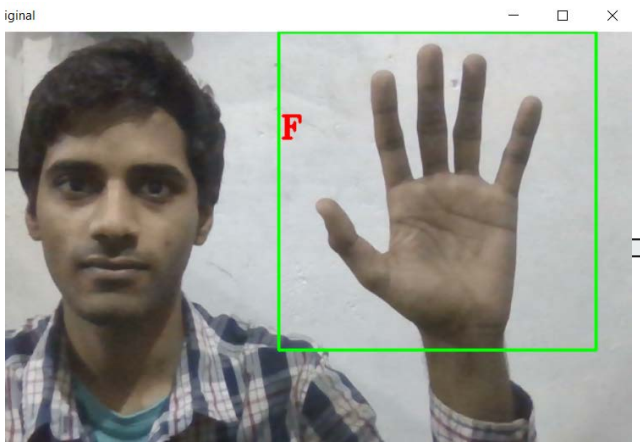
My project aims to create a computer program and train a model which when shown a real time video of hand gestures of American Sign Language shows the output for that particular sign in text format on the screen.

Implemented 26 symbols (A-Z) of ASL in our project.



Methodology

Capturing Raw Image



Cropping the required Area

Image BGR2HSV

Prediction 1

Classify between
26 Symbols

Prediction 2

Classify between
Similar Symbols

Prediction 1:

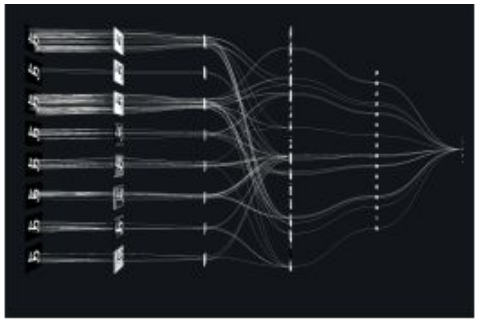
1. Apply BGR2HSV to the frame taken with opencv.
2. This processed image is passed to the CNN model for prediction.
3. After Prediction, if the letter predicted lies inside the Similiar symbol group, then 2nd Model is applied according to requirement and then final answer is printed on the screen with frame.

Prediction 2:

- We detect various sets of symbols which show similar results on getting detected.
- 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 V : K
  2. For W : F
  3. For X : P and F
  4. For Y :G and F
  5. For Z :P and K

Convolutional Neural Networks

- CNNs consist of multiple convolutional layers with each layer containing numerous “filters” which perform feature extraction.
- By training, the feature extraction gets better and better.
- It’s primarily used for image classification.



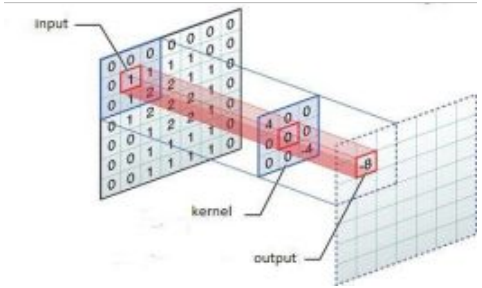
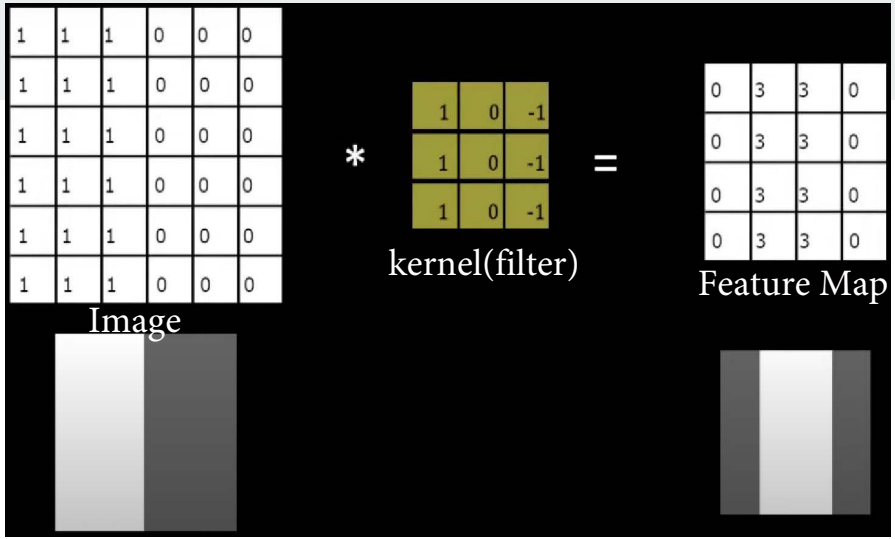
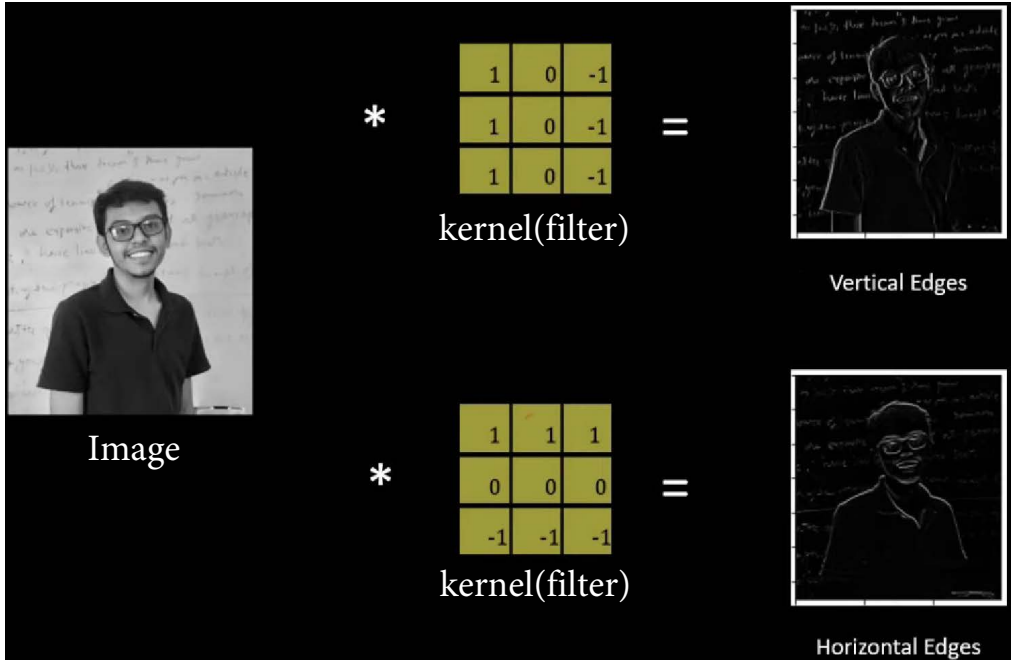


Image * kernel(filter) = Feature Map



Filters extracts specific features from images and feed them into the fully connected layers



CNN Classifier Model

CNN Classifier Model

32×50×50
- Convolution
- Max Pooling

64×50×50
- Convolution
- Max Pooling

64×50×50
- Convolution
- Max Pooling

Flatten

Fully Connected Layer

Output

Challenges Faced

➢ First issue was to select a input format for feature extraction. Tried various formats including RGB, Gray_Scale and HSV(Hue, Saturation, Brightness) of which HSV was giving better results.

➢ Issues were faced relating to the accuracy of the model we trained in earlier phases which we eventually improved by using 2 levels of Predictions.

Software Requirements

- Python 3.6.6
- Tensorflow 1.11.0
- OpenCV 3.4.3.18
- NumPy 1.15.3
- Keras 2.2.1

Limitations of our model

- The model works well only in good lighting conditions.

- Plain background is needed for the model to detect with accuracy.

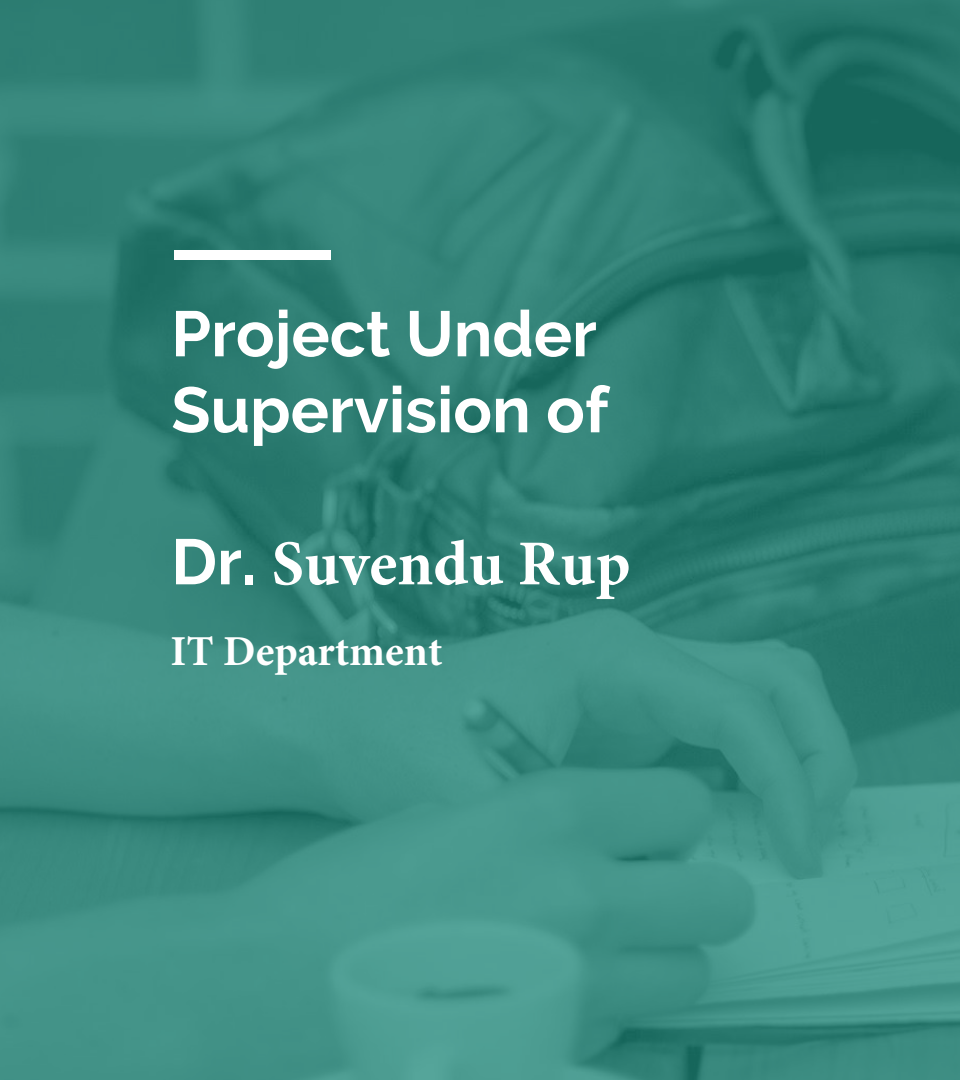
Conclusion

- In this report, a functional real-time vision-based american sign language recognition have been developed for ASL alphabets.
- Achieved an accuracy of **89.7%** 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.

Future Scope

- I am planing to achieve higher accuracy even in case of complex background and low light conditions.

- I am also thinking of adding feature for making sentences. So that it can be used in actual conversation.



Project Under Supervision of

Dr. Suvendu Rup

IT Department

Efforts by

Sumit Soni

Roll No. - 21116104

Branch - Electronics and Communication Engineering

Presently in - 4th Sem, NIT Raipur

Thank You !