Sign-to-Text

By Divya Poluri, Muskan Srivastava, Madhulika P

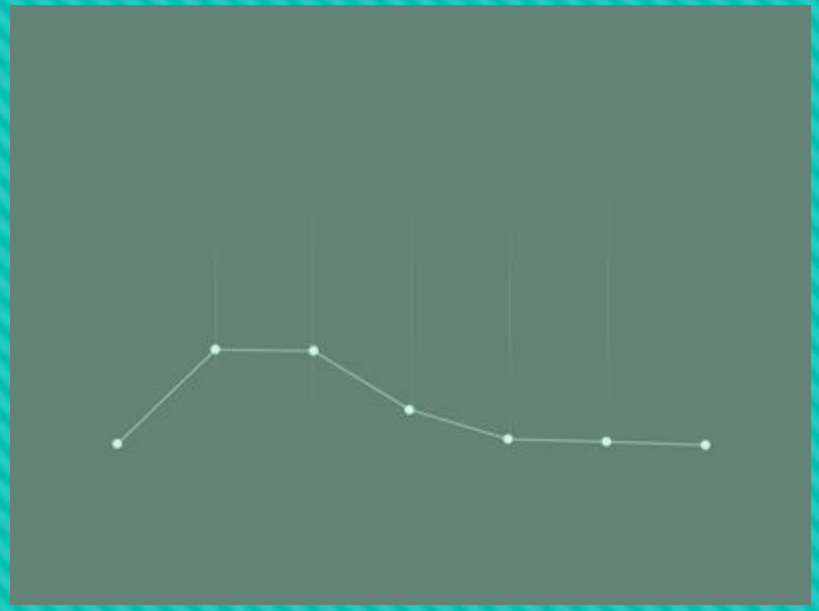
Guided by Prof. Shruti Garg

Motivation

- A person with speaking disorders faces a lot of difficulties expressing their emotions freely in this world. Sign Language is the most natural and expressive way for hearing impaired people. People, who are not deaf, never try to learn the sign language for interacting with the deaf people. This leads to isolation of the deaf people.

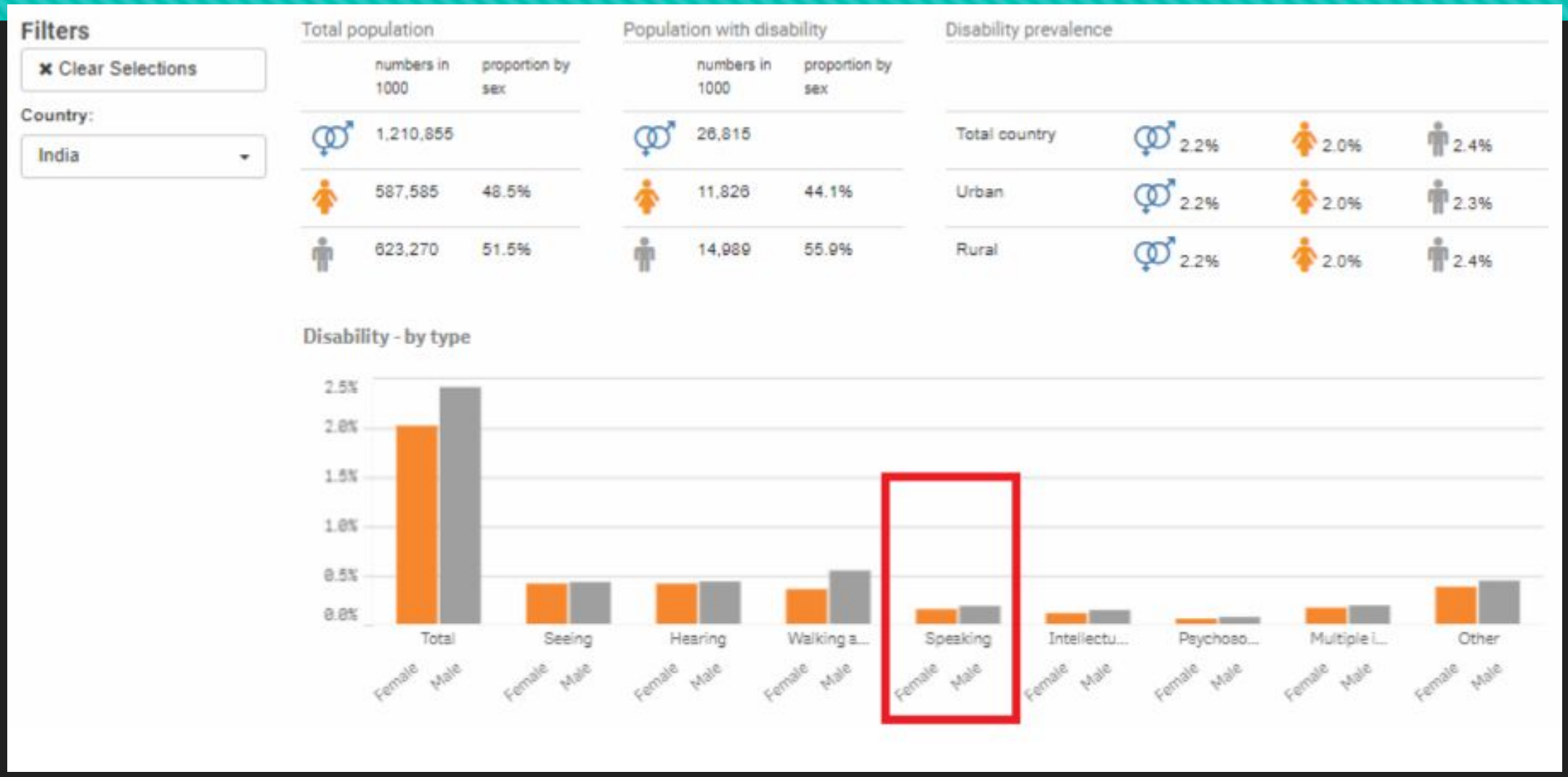
Idea

- Sign Language Recognition is one of the most growing fields of research area. Many new techniques have been developed recently in this area. The Sign Language is mainly used for communication of deaf-dumb people. This project was motivated by our desire to investigate the image recognition field of machine learning since it allows to approach computer vision which is in trend.

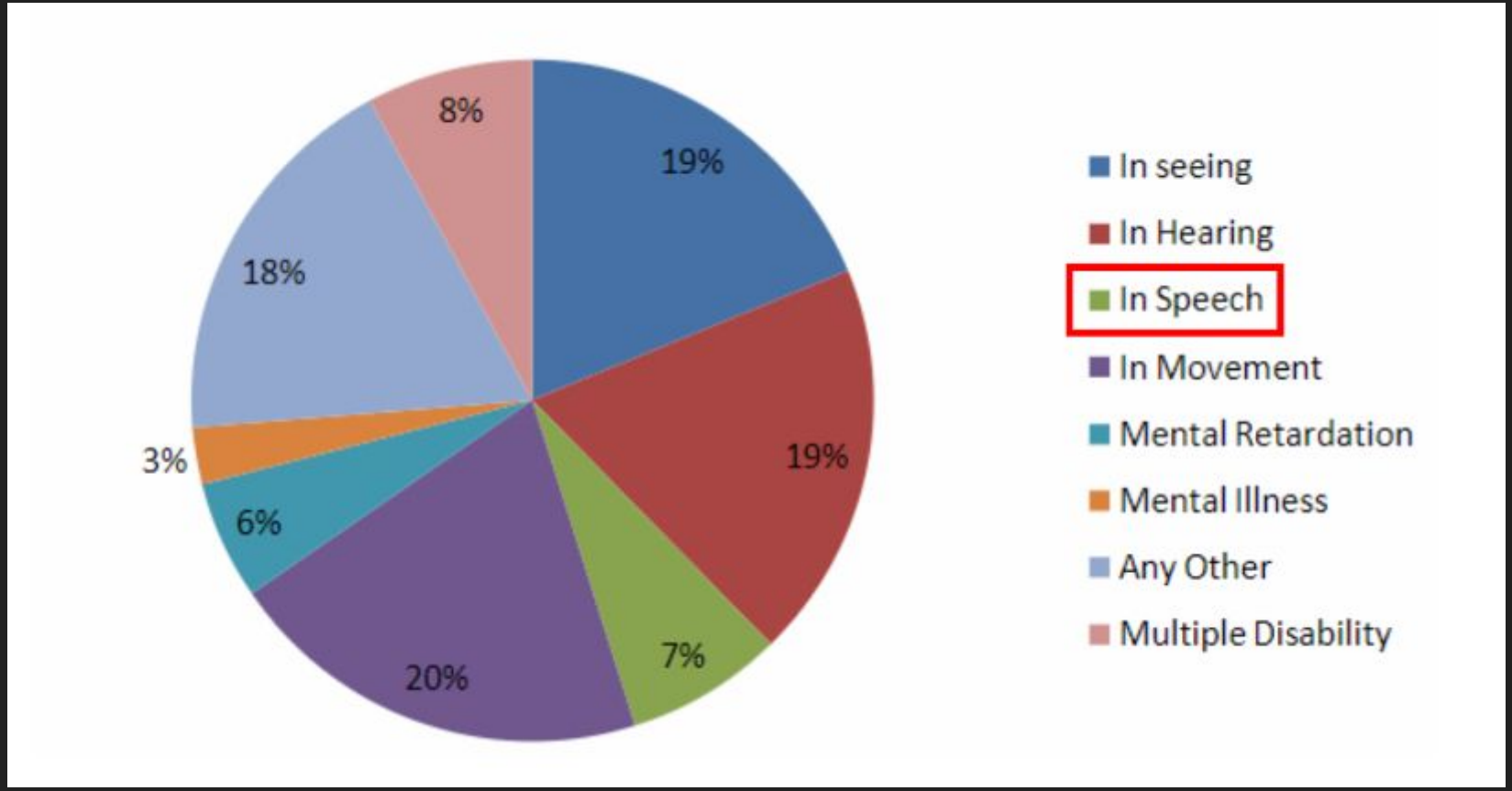


Here are some statistics

Statistics from United Nations

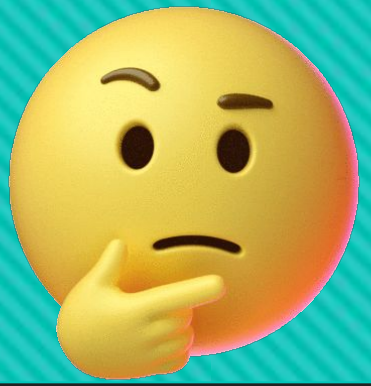


Disabled population by type of disability in India census 2011



- According to the 2011 census, 25% of the population comprises of people who have hearing and speaking disabilities, and communicate using their hands. The problem, however, is that Sign Language is not very popular among the remaining 75% of the Indian population.

Problem statement



- The objective is to build a recognizer that takes an image as an input (or capture images from a camera), recognizes a hand in the image, and classifies the sign (if there is one) into 26 letters in the English alphabet.

Technologies Used

- Python 3.6
- OpenCV
- Deep learning and object detection in images

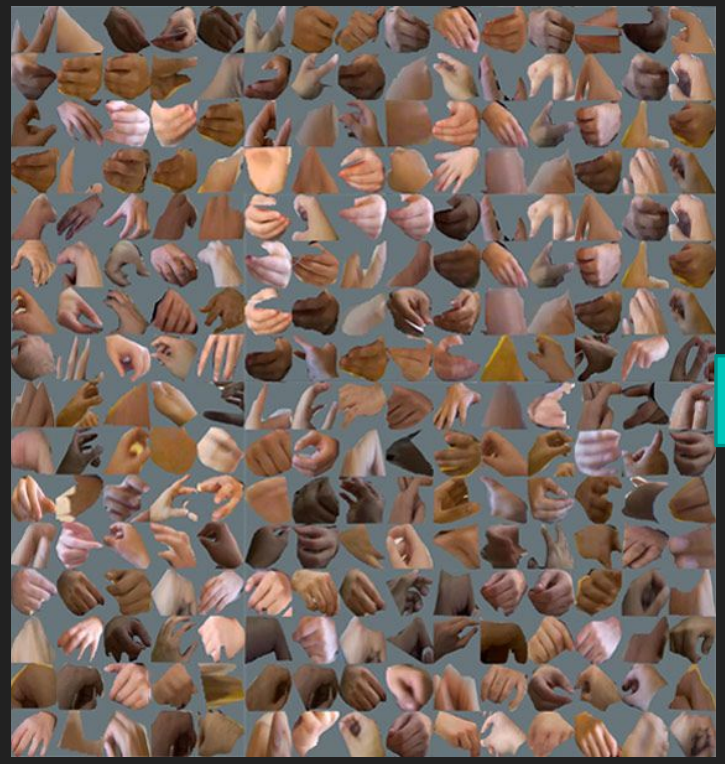
Plan 1.0

1

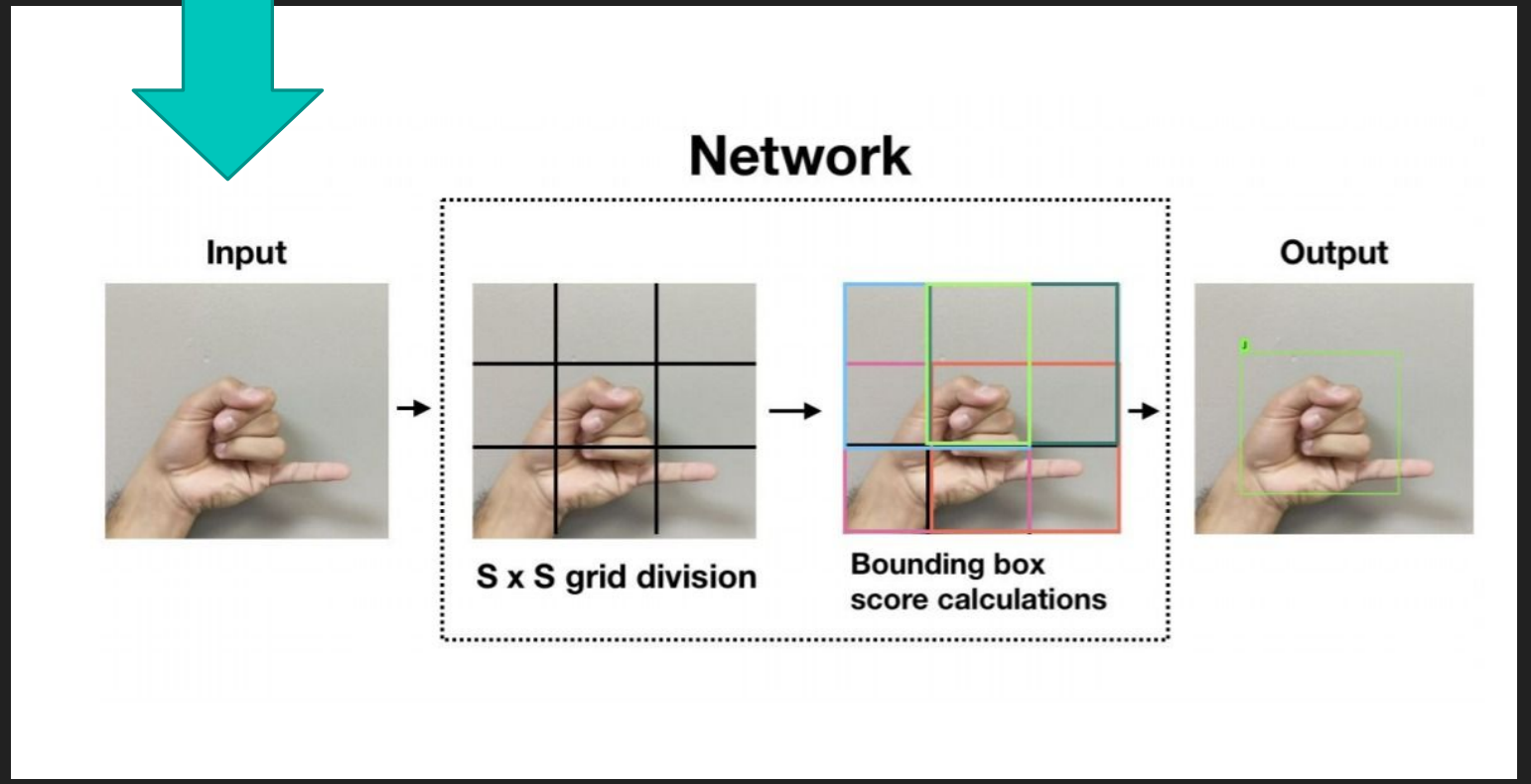
- Use a pre-existing data set, and train it on You Only Look Once for detecting hands and drawing bounding boxes

2

- Extract the image covered in bounded box and feed it to a simpler NN that classifies the hand gesture into one of the 26 alphabets.



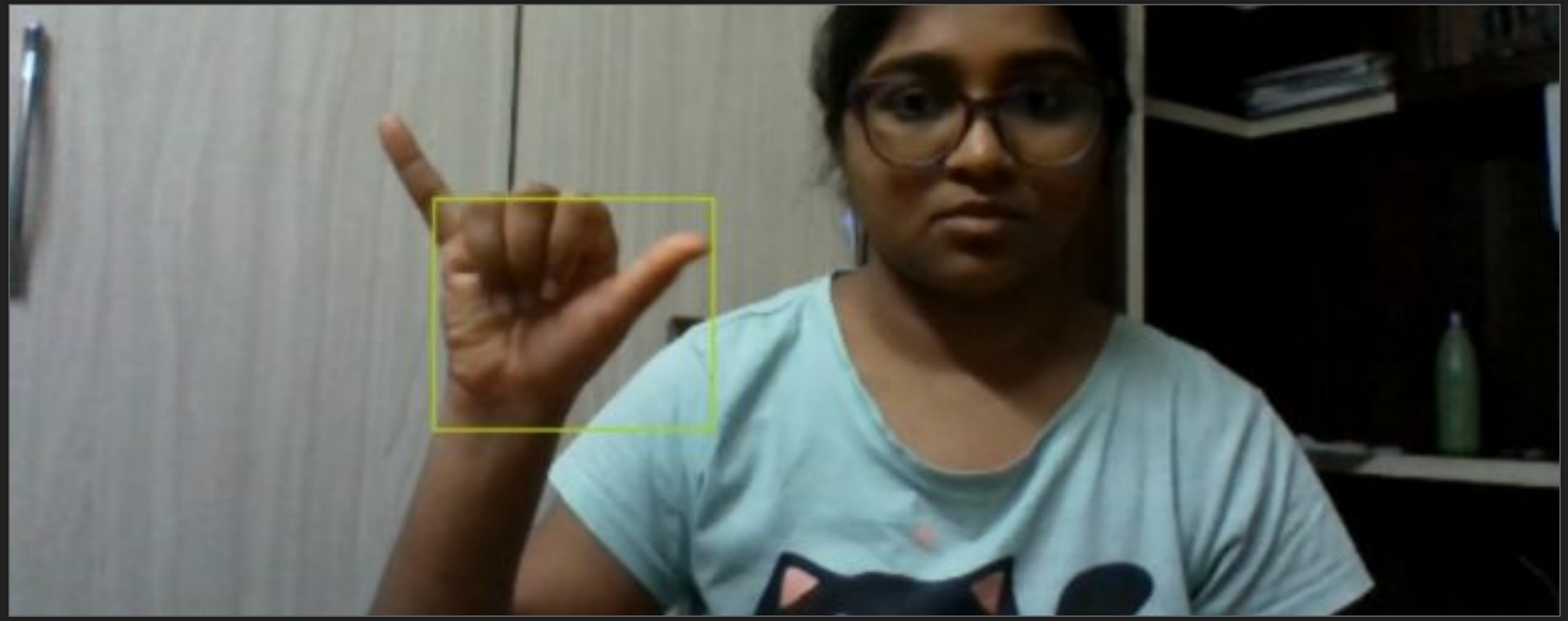
Egohands data set



We faced some unexpected challenges

Challenges Faced

- Training images have a lot of images annotating back of hand
- Fails for certain letters like d
- Inaccurate Bounding Boxes



Inaccurate bounding box

Conclusions

- This approach did not align with our requirements
- Hence, we decided to build our own dataset with annotated images.

Plan 2.0

1. Build a diverse dataset with at least 4000 annotated images
2. Use You Only Look Once (yolo) to recognize gestures

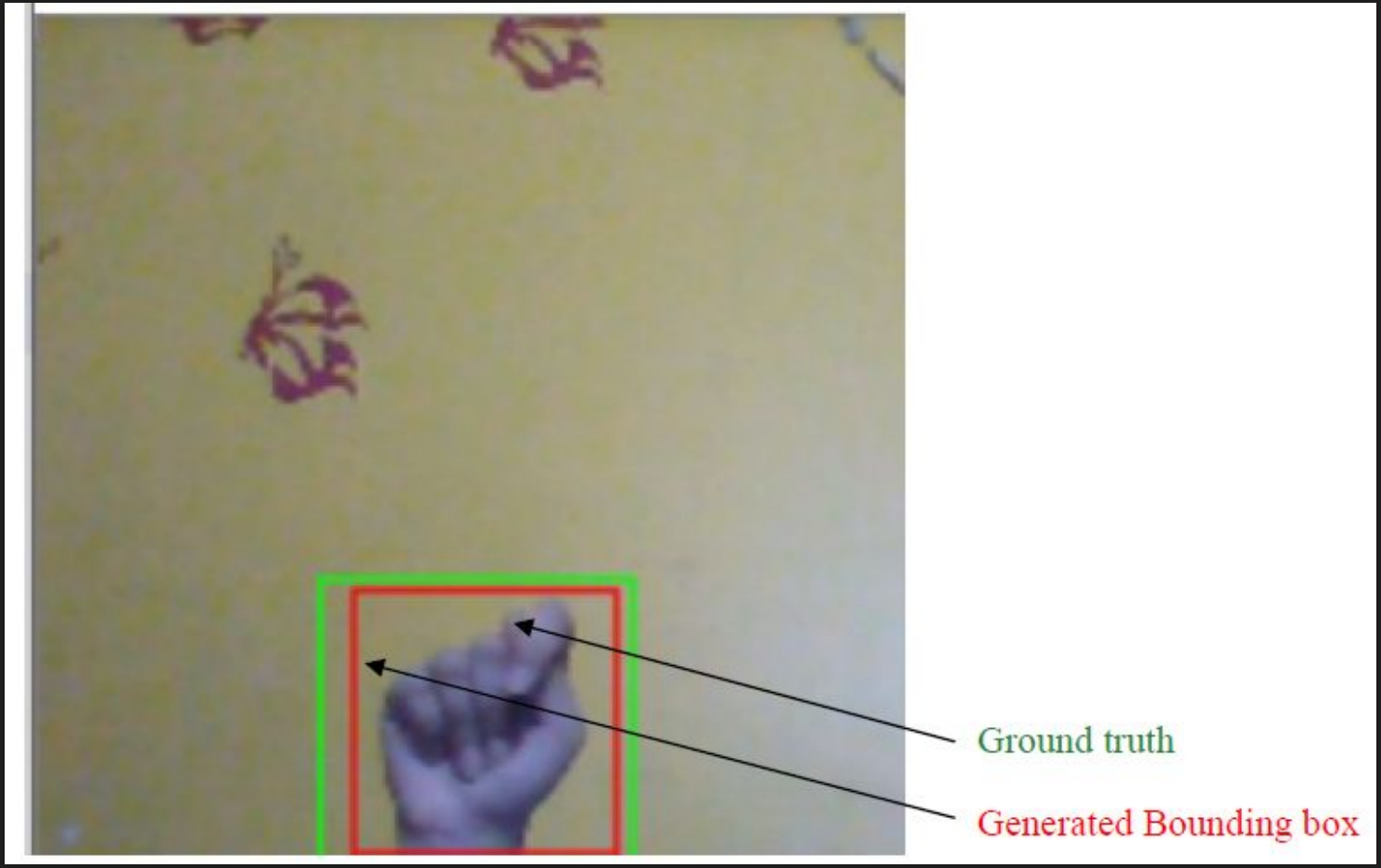
Methodology

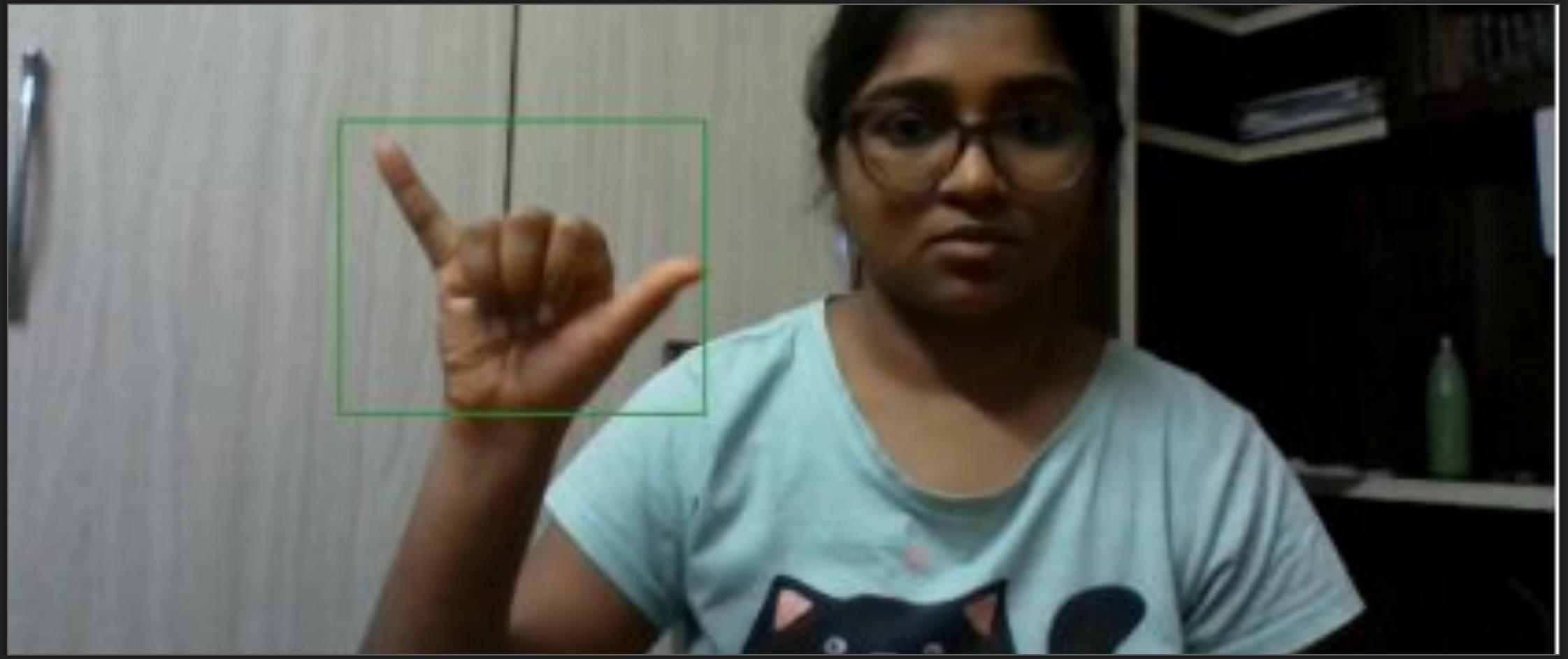
- Build a diverse dataset - Our custom dataset consists of 4,338 images which were split into 80:20 training: test data (3561 training images and 777 test images). These images were annotated using LabelImg. These annotations consist of 27 classes (0:hand, 1:A, 2:B, …, 25:Y, 26:Z). It should be noted here that the letters I and Z require motion of hands, and thus our recognizer was not trained to recognize those two letters.
- Use You Only Look Once V3 (yolo v3) to detect objects- We opted to have only one neural net(based on YOLO) that would directly detect a sign in the image. Training a neural network from scratch is highly expensive and time-consuming, so we used the darknet53 architecture and the pre-trained weights. Using transfer learning, we trained this architecture to detect signs from our labelled dataset.

Why YOLO?

- Other Object Detection networks such as Regional CNNs were not chosen because they take more time to detect objects as they look through an image twice. We wanted to build a model that could further be used for real-time object detection, and networks that look at the image only once seemed a better choice.

Accurate bounding boxes

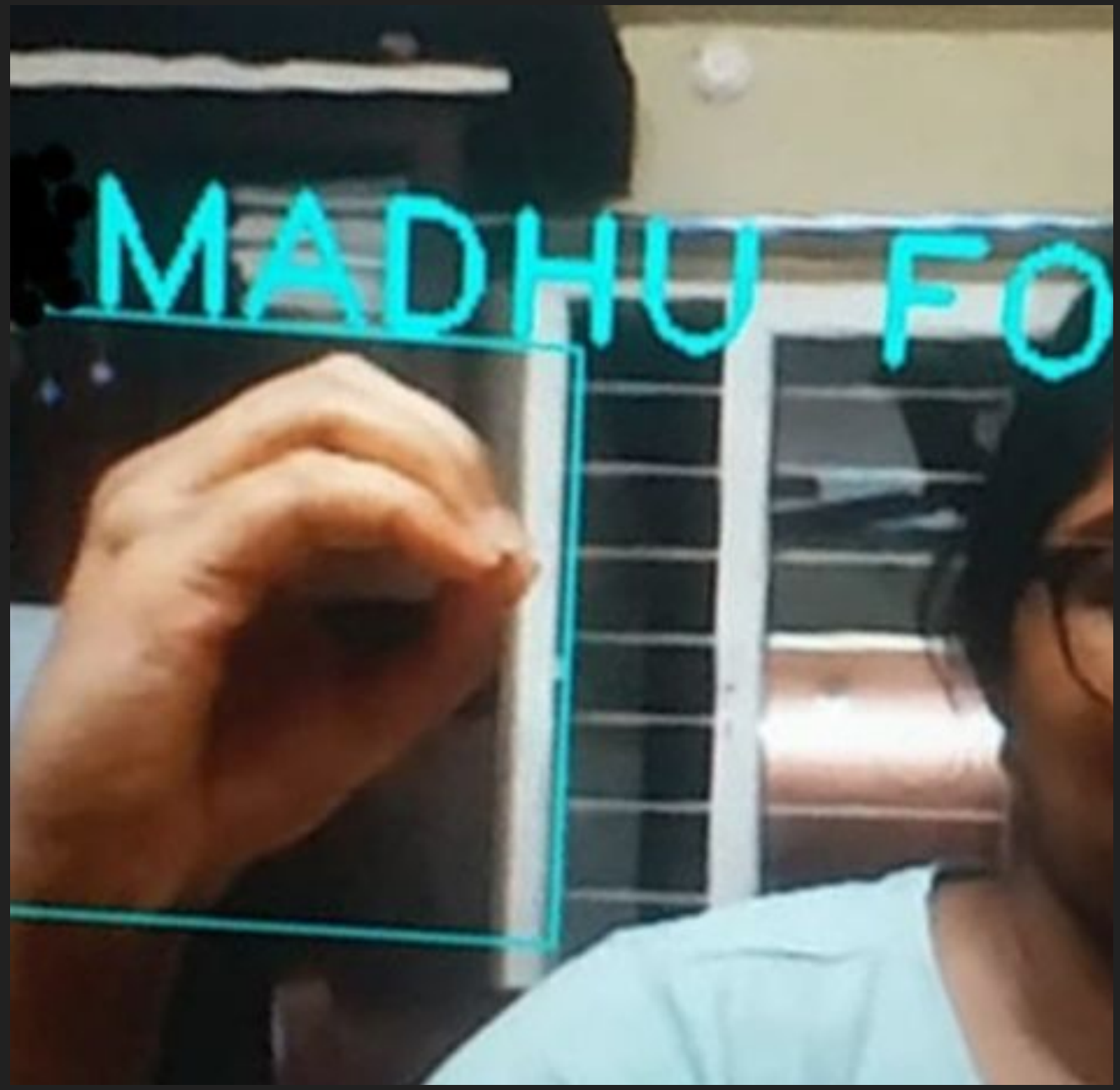




Accurate bounding box

Result

We have successfully built an image recognizing model to recognize American Sign Language(ASL) and convert it to text. Our model has achieved a mAP (mean average precision) of 60.6.

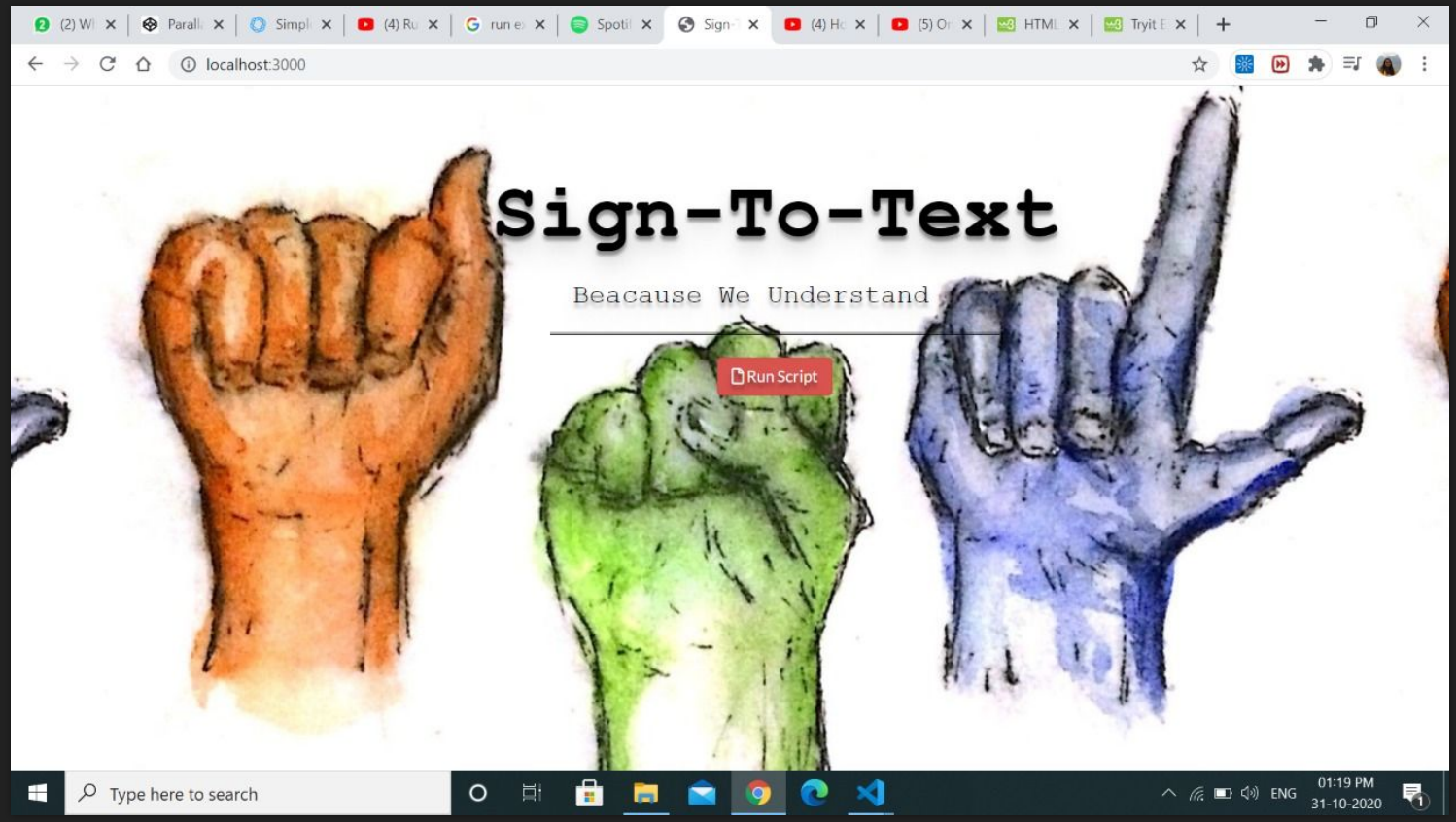


Conclusion

- By using You Only Look Once (YOLO) method, a fast recognizer was built that detects signs in the American Sign Language. This recognizer can be used for real-time hand detection that can be used by people with and without hearing and speech disabilities to communicate effectively with each other.
- We understand that communication is a basic human need and hence, we believe that every individual, despite their disabilities, should be able to communicate without any difficulties. Every God creature has an importance in the society, remembering this fact, let us try to include hearing impaired people in our day-to-day life and live together.

Future Scope

We are planning to embed the recognizer into a website so that more people will be able to access the recognizer easily, thereby helping thousands of people around the world. We aim to create a website using NodeJs and JavaScript features.




Thank
you