SpellVision 2.0

Nicholas Emerson, Ethan Taylor, Carlos Maldonado
Ting Xia

Electrical Engineering & Mechanical Engineering

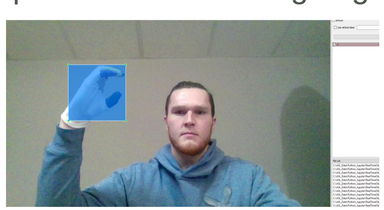
NORTHERN ILLINOIS UNIVERSITY
College of Engineering and Engineering Technology

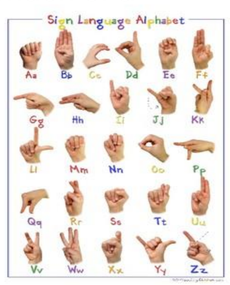
Abstract

This project was created intending to use computer vision to be able to recognize the American Sign Language (ASL) alphabet in real time with high accuracy. The reason for such is the gap between those who can hear perfectly fine and those hard of hearing or even deaf. This can be overcome by taking a large dataset of images that correspond to the letters of the ASL alphabet. These images are labeled according to the letter being signed. They are processed through a neural network using transfer learning to help the machine “learn” what is being signed.

Introduction

A small portion of the population knows how to use sign language. So, for those who are hard of hearing it can be difficult to communicate in many situations. There have been devices to try to fix this issue, but nothing has stood out as a solution for this problem. Using a single webcam and a non-invasive glove seems like a possible solution that won’t be much of a problem on those signing.



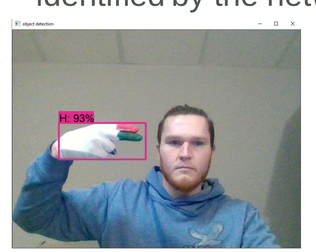


Methods and Materials

Dataset created consisting of 1,440 images including 60 images of each static letter. Each of the images is labeled according to the letter being signed. Image augmentation is done to each image to expand the dataset for the network being trained. The images are separated into training and testing folders. TensorFlow's object detection api and single shot detector (SSD) therein are trained and then ran through a Jupyter Notebook.

Results

The images are successfully trained and tested. The number of training steps used is 20,000 to create high accuracy, but at a cost to time. Without a dedicated graphics processing unit (GPU) the time to train a network takes about 24 hours. The network is able to be used in real time to identify ASL letters after being trained. All designated letters are able to be identified by the network.




```
['N']
['N', 'B']
['N', 'B', 'H']
['N', 'B', 'H', 'E']
['N', 'B', 'H', 'E', 'L']
['N', 'B', 'H', 'E', 'L', 'L']
['N', 'B', 'H', 'E', 'L', 'L', 'O']
['N', 'B', 'H', 'E', 'L', 'L', 'O', 'W']
['N', 'B', 'H', 'E', 'L', 'L', 'O', 'W', 'O']
['N', 'B', 'H', 'E', 'L', 'L', 'O', 'W', 'O', 'R']
['N', 'B', 'H', 'E', 'L', 'L', 'O', 'W', 'O', 'R', 'L']
['N', 'B', 'H', 'E', 'L', 'L', 'O', 'W', 'O', 'R', 'L', 'D']
```

Discussion

Accuracy can be increased with a larger and more diverse dataset. It is important to train the new networks with similar gloves to what was used originally. Using a GPU can help decrease the time needed to train the networks significantly.

Conclusions

This project has the capability of identifying the ASL letters that are static and can be extrapolated to identify whole words easily. Future designs could increase accuracy even more and use a graphical user interface (GUI) to help state what is trying to be said in ASL in sentence format.

Acknowledgements

The authors would like to thank Dr. Ting Xia with the Department of Mechanical Engineering at Northern Illinois University. We would also like to thank Ian Gilmour for being our teaching assistant throughout the entire process.