Real Time Detection of Hand Gestures – With Applications to American Sign Language Alphabet Recognition

Scott Mathews, Satendra Varma, Shyam Narasimhan

Abstract

We present a method for super real-time classification of American Sign Language alphabetical gestures. We employ a two phase image recognition pipeline, wherein hands are localized within the original input image, extracted, and fed into a CNN based classifier which outputs the detected letter. Our results indicate that our method achieves X% accuracy in a real-world scenario.

Introduction

As the power of machine learning techniques continues to grow, the number of interfaces between machine and man continue to grow. For example, assistants such as Alexa and Cortana have popularized a natural speech based interface with computers.
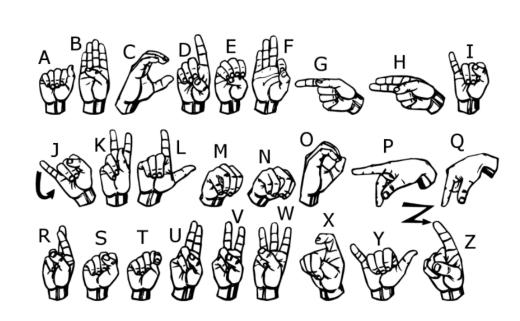
Our goal with this project is to prototype a gesture based interface with computers, with the goal of being able to recognize a predefined set of hand based gestures from a live video feed. To this end, we present an application for recognizing American Sign Language alphabetical hand gestures.

Background and Related Work

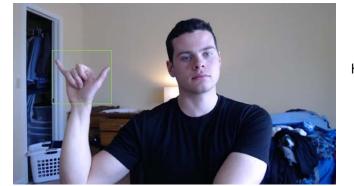Hand localization has been tackled before in the VIVA hand detection challenge.
Sign language letter classification has been approached in a public Kaggle competition titled *Sign Language MNIST*

Methods

Our classifier is trained on American Sign Language alphabetical gestures. Since our model cannot detect movement between frames, letters J and Z have been omitted.

An illustration of our processing pipeline

Hand Localizer → Gesture Classifier → Y

Classifier architecture

Hand Localization

We used a variation of YOLO v2, trained on a dataset of photos with bounding boxes around hands.

Input images from the camera feed are passed into this network, and the bounding box around a hand returned is cropped from the image, then passed to the classifier.

Letter Classification

For classifying the letters from hands, we merged together several datasets of ASL letters. Classification is performed on a 28x28 resized image.

We augmented these datasets with transformed and translated duplicates of images within each dataset.
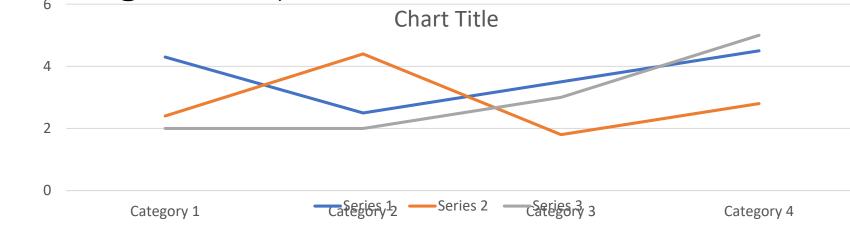
Data

Samples from the localization and classification datsets.

Results

There can be no doubt that our results were stellar. Here's a handy graph of our results. (That is to say, we are still working on them.)

Conclusion

The results clearly demonstrate the superiority of our method compared to other methods for this topic. QED

Actual conclusion forthcoming.