Goals & Motivations

- To use computer vision and machine learning in order to recognize American Sign Language (ASL) gestures in a video recording
- To bridge the gap between American Sign Language users and non-ASL users
- To provide an ASL recognition service as cheap as purchasing a camera, with no additional hardware

Objectives

- **Computer vision**
  - Utilize OpenCV to create a program that uses hand detection as input for data
  - Filter out the background in order to maintain the hand as the only object of focus
- **Machine Learning**
  - Use Keras to model a neural network that can recognize words from the inputs
  - Train the network using over 100 videos per word in ASL to translate the words – “yes”, “no”, “goodbye”, “hey”, and “find”.

Research Challenges

- Finding optimal recording environments
- Finding optimal data formatting to utilize for the Neural Network
- Finding an optimal Neural Network model to utilize for training and translation
- Finding material on video classification due to very niche field of study

Acknowledgement

We would like to thank our advisor, Professor Sarwate, for providing guidance and support.

References


Methodology

- Pass a video through an OpenCV function
- Box out area of the hand and mask background details to simplify input
- Save frames as jpeg files in memory
- Collect 20 frames from each video
- Pass each frame into the Inception-V3 Neural Network to extract features
- Convert the extracted features into a sequence of extracted features
- Use sequence as input to Recurrent Neural Network
- Train model with 1000 samples

Results

- Training our model with 1000 samples, we achieved a **99.60% accuracy** across 16 epochs to classify 5 words
- To provide an even more robust estimate of our model's performance, we evaluated our model using k-fold cross validation. With k = 10 models, we achieved a mean accuracy of **91.67%**.