The idea for this project is to develop an app that can capture a video of a person signaling ASL Alphabet signs and recognize and display correct alphabet sign classes.

The project would consists of two major steps.

The first step should involve successfully identifying the position of both hands in the video image, process it and crop out only hand palms. There are open source solutions and libraries that may guide you in this step.

Once the hand palm is identified and cropped, the second step involves applying CNN like machine learning models in automatically inferring the type of the ASL Alphabet signs being displayed. The model should be trained using based on ASL Alphabet signs dataset at https://www.kaggle.com/grassknoted/asl-alphabet. Whole pipeline should not involve manual process, should flow end-to-end.

Task list

1. Use assignment 1 to collect data on all 26 alphabets by each person in the group (4 tasks)
2. Develop a palm detection algorithm
   a. Use Posenet to obtain wrist points
   b. Develop a cropping algorithm from empirical understanding of the video data
   c. Validate Palm detection algorithm on the videos collected by you
3. Configure a 3D CNN that is trained on the ASL alphabet signs data set in the given link
4. Use your own videos to recognize which alphabets were signed
5. Report accuracy F1 score metrics
6. Now take 40 different words (each student considers 10 different words) and use the alphabet signs to fingerspell these words.
   a. Shoot videos of each of these words using Assignment 1 (4 tasks)
7. Use Posenet on each of these videos to develop a keypoint time series
8. Develop a segmentation algorithm that separates each alphabet in the word
9. Use the same 3D CNN you built to recognize each alphabet in the word
10. Develop an algorithm that combines each alphabet recognition result to develop a recognition of a word
11. Report the word recognition accuracy