1. Spending time
    a. One time video (all signs in one video, at least 2 second between each sign): 4:12
    b. One video for each sign: 9:32 min if using computer
2. Filming person:
    a. 20-30, one person needs to have 2 clips for one sign: one in a light environment and one in dark.
    b. UCSD 1cx class students. I could ask for permission to post an announcement on the blackboard in the ASL classroom.
    c. Spending time is approximately 20 min if film all in two videos and 35 min if film individual ones
    d. 10-15 dollars payment?
3. **Current problems:**
    a. Should I add more signs? The filming process is quicker than I thought. However, running an algorithm on a huge data set may not work.
    b. Do I need to ask the people who film videos to sign some paperworks saying that they acknowledge that their videos will be used in my research and could be open source on the internet for research use in the future.
    c. What is the appropriate payment?
    d. After reading some papers on deep learning of videos. I came up with two possible methods that may work for sign recognition:
        i. Two-stream CNN:
            1. The one I talked about in my proposal.
            2. Pros: Do not need much preprocessing work. Accuracy may be higher.
            3. Cons: Could only work for those signs that in the training data sets. The model training time may be huge. Could only work on a small data set.
        ii. Combination of frames:
            1. Divide one clip to multiple frames and do classification training on each frame based on the hand shape and hand location. Combining the results of each frame and running another classification training. For example, we take 5 frames uniformly from the clip of the sign “good”. Classifying each frame to a combination of a hand shape and a location. Saying the result of the first frame could be (‘b shape’, ‘chin’) and the last frame could be (‘b horizontal shape’, ‘higher chest’). Then running a classification on the combination of 5 results.
            2. Pros: Could be used for signs that may not even be in the training set. Running time could be much shorter since we are not extracting information from the variation between each frame. We are literally dealing with images. Also, data size could be much bigger.

3. Cons: Need a lot of linguistic work before training the model. The accuracy may not be as good as the first method's.