Project Proposal for CS 175, Spring 2018

Project Title: Hand Pose Estimation

List of Team Members:

Christian Shenk, 17803342, cshenk@uci.edu

Jose Eduardo Corona Espinoza, 55984012, coronaje@uci.edu

David Joel Aldarondo, 50177475, daldaron@uci.edu

1. Project Summary

This project will utilize a convolutional neural network to estimate the locations of joints of a hand within an image. The annotation data gathered in class will function as both our training and test data. We will evaluate our model based on its classification accuracy.

2. Problem Definition

Our goal to take a still image or a sequence of still images, and, time permitting, video of hands, and estimate the location of joints of those hands. Specifically, our network would take a still image of a hand as input, then reproduce the photo with nodes representing joints, and with lines between those nodes representing the connecting bones. While the stated goal seems simple, factors like hand occlusion and blurring complicate prediction and training, as it would be difficult for the model to make a guess as to where the joints are.

3. Proposed Technical Approach

We believe using a multi-layered convolutional neural network is an effective solution to our proposed problem. Each kernel will enable us to extract specific features like curvature, different fingers, and change in angle, etc., from each frame of our class data. Our CNN will likely have an architecture along the lines of this: affine transformation -> leaky ReLU non-linear transformation -> Max pool -> affine -> leaky ReLU -> affine -> softmax, similar to the architecture of the CNN of the homework. We would mostly base our algorithms on those discussed in class should we implement the network from scratch. However, we may use the ResNet framework as our source code.

4. Data Sets

We will use the data provided from class as our initial data set, which should be around a few thousand images given that each student has about 800 images associated with them. Should our model perform well with it, and we have enough time, we may move to using recorded videos. Then, should that go well, and, again, time permitting, we will use live video. The increased complexity resulting from moving from still frames to video will test the robustness of our model.

5. Experiments and Evaluation

We will partition the data we receive from class and use cross validation to train our model. Our model's performance will be evaluated based on its predictive accuracy, i.e. whether it correctly identifies joints, even when they are occluded.

6. Software

Based on what other students have suggested, we may use and modify the ResNet framework to build a CNN to estimate hand poses. However, should it not suit our needs, we will instead build our own network using the PyTorch framework initially. We may branch out to Keras or TensorFlow should they suit our needs better.

7. Individual Student Responsibilities

Christian Shenk:

- Implementing the CNN's framework.
- Generating data should we decide to use video.

Eduardo Corona:

- Implementing the CNN's framework.
- Tune hyperparameters.
- Convert CNN to CoreML format if time allows
- Generating data should we decide to use video

David Aldarondo: Has the most powerful computer

- Train model

- implement the cross validation algorithm
- Tune hyperparameters.
- Generate data should we decide to use video