Hand Gesture Recognition

Objectives

UNIMODAL INPUT

Using a single image as an input to perform Hand Gesture recognition. Will allow for higher portability and ease of use

PROVE CORRELATION BETWEEN HAND GESTURE TASKS

Show that the learning from HAND POSE ESTIMATION can simplify the HAND GESTURE RECOGNITION task. Implement classes for performing pose-to-gesture classification, using Random Forests and K-means clustering.

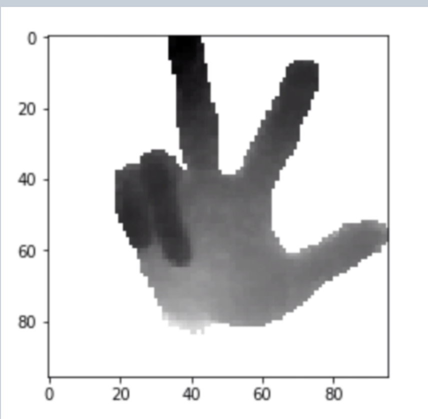
Related Works

DEEP PRIOR++ The latent space of joints is not the complete 3D space. Employ a pinch layer to enforce a low dimensional embedding into the network.

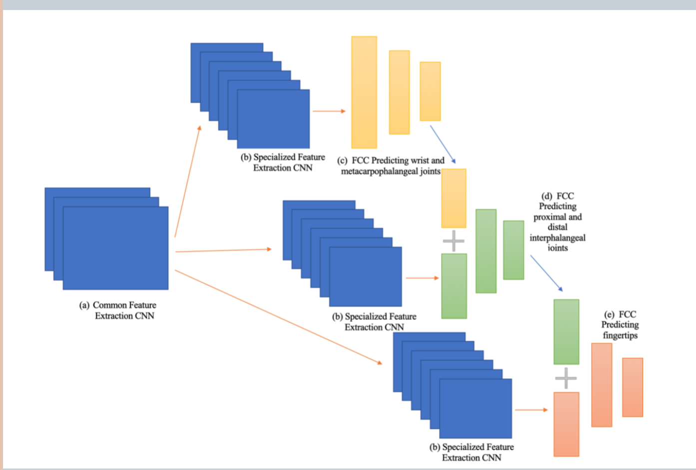
CONVOLUTION POSE MACHINES: Uses the positions of the larger body parts to predict the locations of the smaller body parts.

CROSSINFONET Divides the task of predicting the hand joints into two sub tasks: 1) Palm joint 2) Finger joints. The theorize that the noise for one task is valuable to the other task. Use information sharing to exchange information.

Methodology



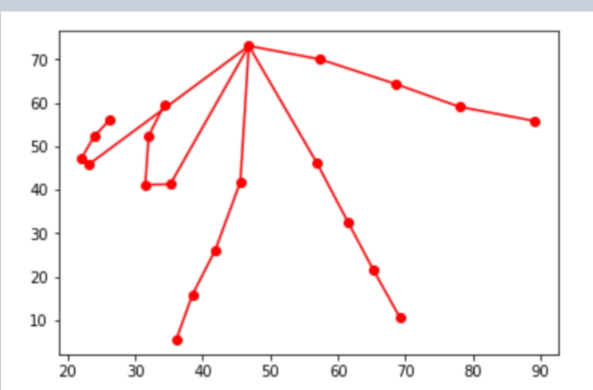
We used the depth images in the MSRA dataset as our main dataset. It contains samples collected from 9 different subjects with annotations for 17 different gestures. On average, there are 500 frames per gesture per person.



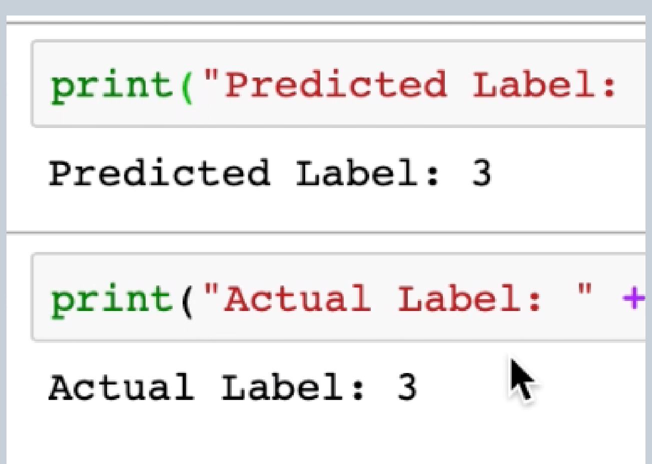
1)The input image is first processed by the common feature extraction layer.

2) The output is then processed by special feature extraction layer which forwards the output to FCC1. The FCC1 predicts the position of the palm joints and forwards its output to FCC 2.

3) This process repeats two more times to give the full pose vector.

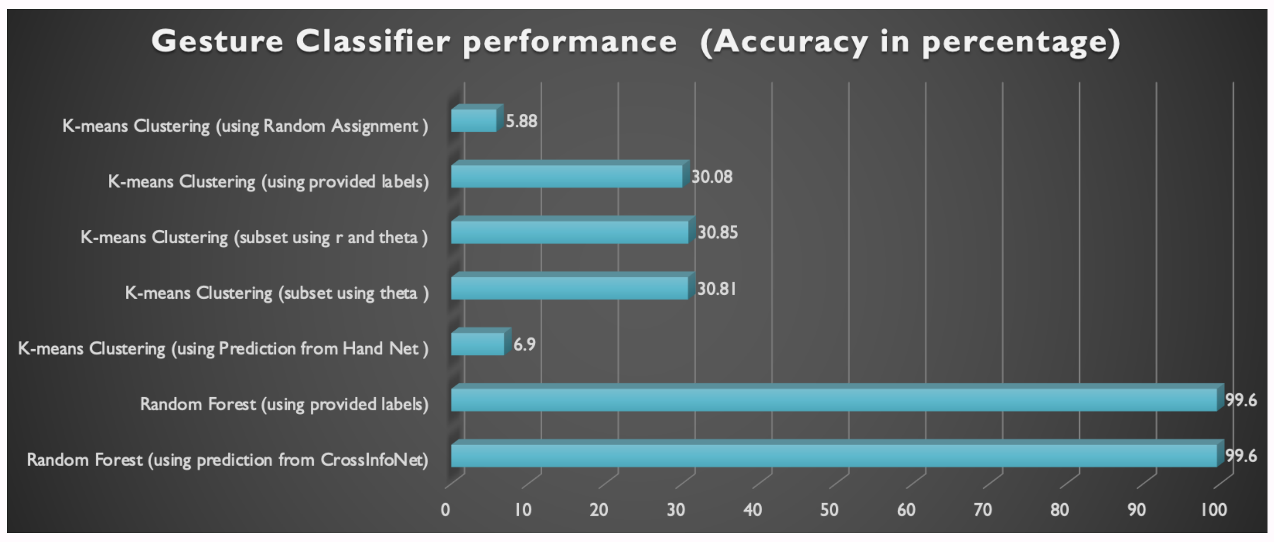


Using r,θ,z coordinate system. Highlights the structure of the hand more naturally. Similar r, θ,z coordinates for x,y,z, and u,v,d coordinates. Allows for assigning weights to different joints. Allows selecting the subsets of coordinates more naturally.



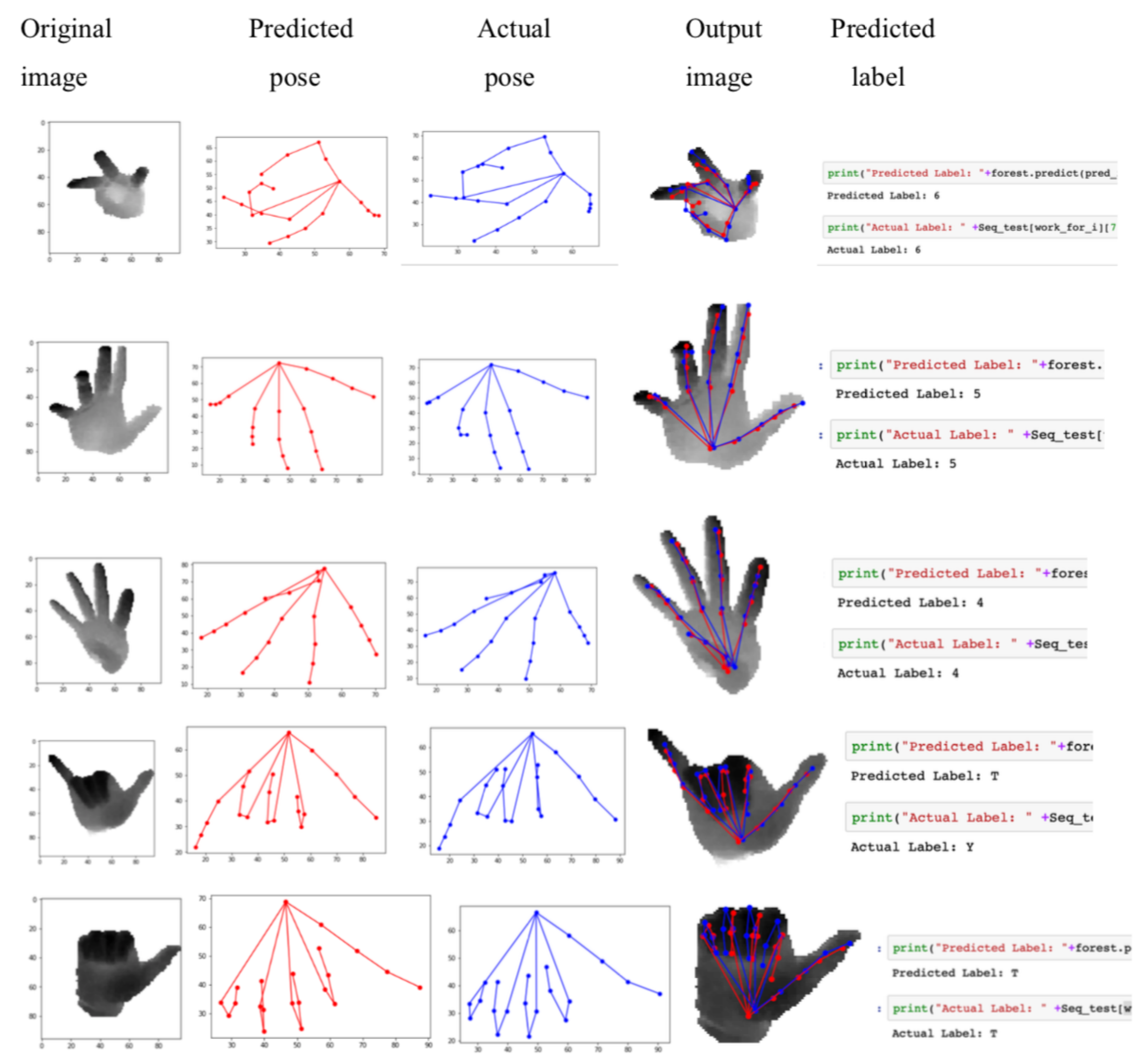
Since pose estimation locates actual points in the image, keypoint tracking can aid pose estimators in a real time application. Therefore, when performing pose-to-gesture learning, all deep learning tasks are left to the hand pose estimators. Therefore, we can leverage systems such a key point tracking for classification. Lastly, the pose-to-gesture model can use models such as Random Forests and execute very quickly.

Results



We observe state of the art performance when using a Combination of CrossInfoNet with Random Forest.

Demonstration



Dhruv Agrawal
Supervisor: Dr. K.K.Y. Wong
Second Examiner: Dr. A.T.C. Tam