# Hand Action Recognition in G-D Video Data

Luke Jaffe

03/18/2018

# Introduction

- **Problem:** Learn real-time hand action recognition for the Pico Monstar camera, which produces Grayscale+Depth data with infrared sensor
- FS Studio, a local Software company, generously let me borrow their Pico Monstar camera
- Can we learn hand action recognition from a proxy dataset?
- Can we learn hand action recognition from label-starved data?
- Can our architecture be adapted to the real-time scenario?

# Approach

- 3D Convolutional Neural Networks (convolve over time dimension in addition to 2 spatial dimensions)
- Approach from paper: Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?
- Updated their architecture for better real-time performance by training CNN features with LSTM

# Proxy Dataset

- Can we learn hand action recognition from a proxy dataset?
- SKIG dataset as proxy, has 1,080 videos of 10 hand actions under different illumination, background conditions, with different poses
- We recorded each action under 3 illumination conditions, for a total of 30 videos
- Trained network on SKIG, tested on data from Pico Monstar prepared to be very similar
- Result slightly better than random (20%): domain adaptation recommended

# Training Native

- If we do not have a dataset which is a good proxy, we can always make our own dataset
- We train from 20/30 of our videos and test on 10/30
- Test results are good, all 10 test videos correct with score-sum argmax on CNN features, ~85% of individual test frames are correct with LSTM

# Real-Time Pipeline

- Frames are recorded in real time from Pico Monstar with C++ SDK
- Frames are saved to disk individually in .npy format
- Python binding is wrapped around C++
- 0MQ sends messages from from producer to consumer process
- Consumer process takes frames, stacks them in groups, puts them through NN
- Resulting prediction displayed with image