2/5/2019

- Met with Professor and grad student Klas and finalized the focus on our research
- Exploring different interpretation models for combating adversarial attacks on facial recognition systems.
- Prior work done by another research group at CMU, paper found here: https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

2/12/2019

- Additional background readings into CNNs to understand terminology and concepts
- Paper readings on prior work done in the research area
- Test dataset used by earlier research group obtained and our rudimentary framework was added to analyze what was going on in the images
- Began account setup to start exploration and testing

2/26/2019

- Introduced to code repository classes and Jupyter
- Introduced to which sections of code that could be modified for exploration
- Preparations to start exploration process over break

3/19/2019

- Exploration in progress for the next several weeks.
- Created visualization tool by viewing influence across layers
- Created visualization tool contour mapping influence on input layer
- Created visualization tool by viewing multiple thresholds of influence

4/2/2019

Completed heatmaps of influence for 2 batches of tests on 2 different networks for all layers, which seems to confirm a possible heuristic can be detected and used for a classifier. Started looking at feature maps at layers where this phenomenom was most prevalent, and created a characteristic profile (feature vector mean across images). Signal to noise ratio is a bit problematic for comparisons.