# To Post or Not To Post: Using CNNs to Classify Social Media Worthy Images

Lauren Blake (lblake@stanford.edu)

## Summary

**Problem:** Determine the feasibility for CNNs to predict whether photos depict our best selves

**Approach:** Classify images based on their social media worthiness score using a CNN trained with transfer learning

**Evaluation:** Classification accuracy & heat maps

**Key Results:** Classification results are promising with reasonable accuracy rates and heat maps. Yet, the results also highlight differences between the model's and a human's vision

## Background

People are picky about how they are depicted in photographs

Eyes closed
Looking away
Not smiling
Many other concerns!

Numerous exciting applications exist to apply these CNN models to improve photography

**Camera**

- Provides real-time feedback to the photographer
- Knows when to take photos

**Software**

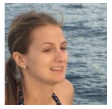
- Automatically posts social media worthy images
- Recommends top photos from an album

**Future Applications**

Relevant prior work classifies other abstract qualities (e.g. beauty, emotion) instead of objects in an image. Best self has not been studied yet

## Dataset

Data set contained 424 images of my face, which were cropped photos from my personal collection. Each image was assigned a social media worthiness score of 1 to 5

| | Score 1 | Score 2 | Score 3 | Score 4 | Score 5 |
|---|---|---|---|---|---|
| |  |  |  |  |  |
| Opinion | Hate | Dislike | Neutral | Like | Love |
| Sharing | Absolutely Not | Avoid | Hesitant | Some | Broad |
| % of Data | 14% | 21% | 33% | 24% | 8% |

**Example Images and Their Social Media Worthiness Scores**

## Methods

Due to small dataset size, took advantage of transfer learning and used a pre-trained CNN as a fixed feature extractor

ResNet-18 Model

*Pre-trained with ImageNet data*

ResNet-18 with new final linear layer

*Final layer parameters trained with our data*

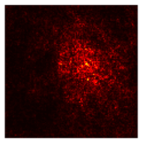
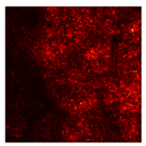
Training Process
- Batch SGD
- Cross entropy loss
- Learning rate & momentum turning

## Findings

Model had over 50% accuracy with particularly strong performance for Score 1 & 3 images

Heat maps support that classification is driven by important facial features (especially eyes). Results are consistent across different poses and scores

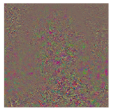
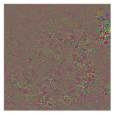
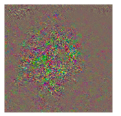
| Actual Score | 1 | 2 | 3 | 4 | 5 | Overall |
|---|---|---|---|---|---|---|
| Classification Accuracy | 67% | 42% | 63% | 42% | 0% | **51%** |
| Mean Abs. Error | 0.3 | 0.7 | 0.5 | 0.8 | 1.5 | **0.6** |

| | Score 1 | Score 3 |
|---|---|---|
| Original |  |  |
| Heat Map |  |  |

## Fooling Images

Fooling images result from altering an image in order to "trick" the model to change its score

Changes that are nearly invisible to human eyes are able to significantly change image scores. People are unlikely to be tricked by these changes

| | Original Image | Score 5 Fooling Image | Magnified Difference (50x) |
|---|---|---|---|
| Score 1 |  |  |  |
| Score 3 |  |  |  |
| Score 4 |  |  |  |

## Limitations

Very small data set with only one person. In particular, the data doesn't include people with different genders, races, or ethnicities

Would ideally train with Euclidean loss to capture the sequential relationship between scores

ResNet-18 model may not be the optimal feature extractor for face images