How to Take a Good Selfie?

Mahdi M. Kalayeh[⋆] Misrak Seifu[◦] Wesna LaLanne[⋄] Mubarak Shah[⋆]
[⋆]Center for Research in Computer Vision at University of Central Florida
[◦]Jackson State University [⋄]University of Central Florida
mahdi@eecs.ucf.edu misrak.seifu@gmail.com wlalanne@knights.ucf.edu
shah@crcv.ucf.edu

ABSTRACT

Selfies are now a global phenomenon. This massive number of self-portrait images taken and shared on social media is revolutionizing the way people introduce themselves and the circle of their friends to the world. While taking photos of oneself can be seen simply as recording personal memories, the urge to share them with other people adds an exclusive sensation to the selfies. Due to the Big Data nature of selfies, it is nearly impossible to analyze them manually. In this paper, we provide, to the best of our knowledge, the first selfie dataset for research purposes with more than 46,000 images. We address interesting questions about selfies, including how appearance of certain objects, concepts and attributes influences the popularity of selfies. We also study the correlation between popularity and sentiment in selfie images. In a nutshell, from a large scale dataset, we automatically infer what makes a selfie a *good selfie*. We believe that this research creates new opportunities for social, psychological and behavioral scientists to study selfies from a large scale point of view, a perspective that best fits the nature of the selfie phenomenon.

Categories and Subject Descriptors: I.4 [Image Processing and Computer Vision]: Applications

Keywords: Selfie; Social Media; Sentiment Analysis; Recommendation Models.

1. INTRODUCTION

According to the Oxford Dictionary, *Selfie is a photograph that one has taken of oneself, typically one taken with a smartphone or webcam and shared via social media.* In the past few years, taking selfies has become very popular. People from different socio-economic, gender, race and age groups take selfies in various occasions. Google recently reported[1] that there are 93 million selfies taken every day only on Android devices. These gigantic data can reveal interesting statistics about the preferences, moods and feelings of the members of our society, if it is properly analyzed. Due to the large scale and continuous growth of the data, it is appropriate to employ machine learning and computer vision techniques to study selfie images. We believe that when someone takes a selfie, he/she considers two major aspects among many possible others to make it a *good selfie*, popularity and sentiment. Popularity refers to the log_2 normalized *view* counts while sentiment indicates the feeling that viewers infer by looking at a selfie. Typically, a social media user desires to increase the popularity of his/her selfie and imply a positive sentiment. In this paper we are addressing the following questions: **(1)** How do different attributes, such as gender, race or hair color, influence the popularity of selfies? **(2)** How does the appearance of certain objects or particular concepts affect the popularity of selfies? **(3)** Is there a relationship between the sentiment inferred from a selfie and its popularity? **(4)** How does post-processing, such as applying different Instagram filters, influence the popularity of selfies?

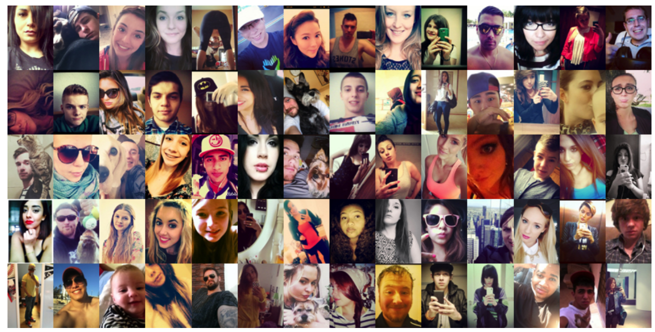

Figure 1: Examples of images in the Selfie dataset.

Recently, Khosla et al. [4] proposed a framework to predict the popularity of a photograph before being uploaded on social media. They use about 2.3 million images from Flickr to predict the normalized number of views for images. Exploiting both visual content and social cues, Khosla et al. predict the popularity of an image with about 0.81 rank correlation to its ground truth. Borth et al. [2] have proposed sentiment prediction based on both visual content and tags associated to the images. They [2] use SentiBank, a library of visual concept detectors based on more than 2,000 Adjective Noun Pairs (ANP) to predict the implied emotion of a photograph. While these studies are valuable, they deal with photographs in general. We believe that studying selfies as the current most popular type of shared content on so-

[1]`http://goo.gl/53ZOjG`

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MM'15, October 26–30, 2015, Brisbane, Australia.
© 2015 ACM. ISBN 978-1-4503-3459-4/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2733373.2806365.

cial media deserves an exclusive study. Knowing that photos with human faces are respectively 38% and 32% more likely to receive *likes* and *comments*, Bakhshi et al. [1], further demonstrate the importance of this work. Our study does not focus on improving the popularity score prediction or boosting the sentiment prediction performance. Instead, we are using state-of-the-art techniques in these areas to simply answer the four aforementioned questions. In the rest of this paper, we introduce our Selfie dataset in section 2. In section 3, we present an attribute prediction baseline. In section 4, we evaluate how does the appearance of different objects and concepts influence the popularity of selfies. In section 5, we answer the question of correlation between popularity and sentiment. Finally in section 6, we study whether applying different Instagram filters affects the popularity of selfies. We conclude the paper in section 7.

2. SELFIE DATASET

Due to the massive number of selfies being uploaded on social media every minute, any study aiming to provide insights about selfies has to be conducted on very large number of images to ensure that it sufficiently captures the variations in the data. To the best of our knowledge, there exists no selfie dataset publicly available for research purposes. Therefore, we collected our own dataset. We downloaded 85,000 images from `selfeed.com`, a real-time update of #*selfie* on Instagram. Despite being tagged with #*selfie*, we found that only 69,710 of those images are actually selfie images. The remaining 15,290 images were either completely irrelevant or general photographs of people. This is an interesting observation, since different social media usually report the number of selfies shared on their environment by counting the images tagged with #*selfie*. However, in our data collection, at least 18% of images tagged with #*selfie* are not, in fact, selfies. In preparation of the Selfie dataset, we made sure to discard any images that do not fall under the definition of *selfie*. Clearly this is a subjective judgment and different annotators may disagree in a few cases whether a photo is a selfie or not. In order to generate a uniform understanding of *selfie* among our annotators, we asked them to annotate a fixed set of images containing about 2,000 images. Then, the images with mixed votes where discussed and disagreement was resolved by clarifying the definition. About 32% of selfie images that we collected were showing multiple faces either in form of group selfies or collages. Excluding multiple-face images yields to a total number of 46,836 single-face selfies. Since we wanted to annotate selfies with attributes such as age, gender, hair color and etc., we were mostly interested in selfies that do not show multiple faces. Figure 1 shows some of the collected selfie images. We annotated 46,836 selfie images with 36 different attributes divided into several categories as follows: *Gender:* is female. *Age:* baby, child, teenager, youth, middle age, senior. *Race:* white, black, asian. *Face shape:* oval, round, heart. *Facial gestures:* smiling, frowning, mouth open, tongue out, duck face. *Hair color:* black, blond, brown, red. *Hair shape:* curly, straight, braid. *Accessories:* glasses, sunglasses, lipstick, hat, earphone. *Misc.:* showing cellphone, using mirror, having braces, partial face. *Lighting condition:* harsh, dim.

We asked annotators to look into a random subset (∼3,000 images) of the Selfie dataset and create a list of attributes that they 1) frequently observe and 2) can easily detect. These two conditions assure enough positive samples for each attribute and an acceptable performance for the attribute detectors. Figure 2 shows the ground truth statistics of the collected Selfie dataset.

3. ATTRIBUTE PREDICTION BASELINE

To analyze selfie images outside our dataset, we have to be able to predict different attributes of interest with an acceptable precision. Therefore, we provide a baseline for attribute prediction along with introduction of our Selfie dataset. In our experimental setting, for every attribute, positive instances were randomly divided into 3 folds. We do the same for the negative instances. Training split for each attribute consists of 2 out of 3 folds from both positive and negative instances. The third folds from positive and negative instances, together create the testing split. For feature extraction from images, we use

SIFT: We densely extract SIFT descriptors from images at every 3 pixels and at 8 different scales. To encode these descriptors into a single feature vector, we employ VLAD [3] with a codebook size of 256.

HOG: We densely extract HOG descriptors from images with the cell size of 8 pixels. Descriptor encoding is the same as the one used for dense SIFT.

Deep Features: Using deep convolutional neural networks (CNN) [5] trained on ImageNet dataset [6], we extract 4096-D feature vectors from CNN's last fully connected layer. We also use final 1000-D classification layer as another feature representation. We employed the large network of OverFeat [7] to implement CNN.

SentiBank: Authors in [2] have trained 2,089 visual concept detectors based on a list of Adjective Noun Pairs (ANP). These ANPs are like *smiling boy*, *lovely dress*, *scary face* and etc. Using SentiBank, we generate a 2089-D vector for each image where every dimension is the detection score of its corresponding ANP concept detector.

We employed *one-vs-all* support vector machine (SVM) with linear kernel to train attribute detectors. The performance of detectors is measured via average precision (AP). Figure 3 shows the performance of different features for the task of attribute detection on the Selfie dataset. Deep features extracted from last fully connected layer of CNN perform better (29.51% vs 24.03% meanAP) than the 1000-D features from CNN's classification layer. Features obtained by applying SentiBank ANP concept detectors achieve the best performance with 31.97% meanAP among three different mid-level features. We expected to observe this since ANPs are very diverse and a very large portion of them are relevant to human attributes. Using SIFT and HOG (low-level features) with VLAD encoding results in 33.76% and 22.95% meanAP, respectively. We also fused 4096-D deep features, SentiBank features and SIFT via mean and max pooling their detection scores (late fusion). While max pooling was not helpful (32.49% meanAP), mean pooling further improved the attribute prediction baseline to 35.79% meanAP.

4. WHAT MAKES A SELFIE POPULAR?

In this section, we attempt to answer the second question that we proposed in the beginning of this paper. How does the appearance of certain objects or particular concepts

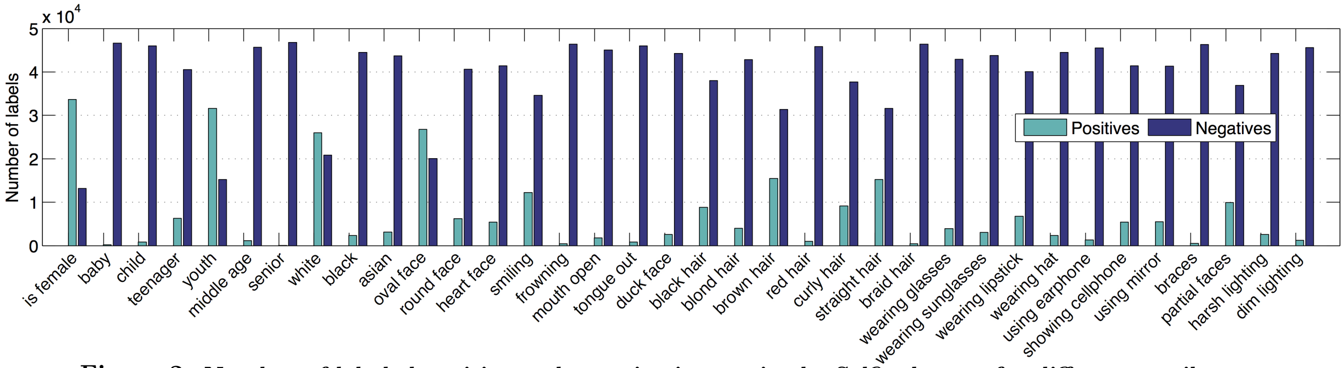

Figure 2: Number of labeled positive and negative images in the Selfie dataset for different attributes.

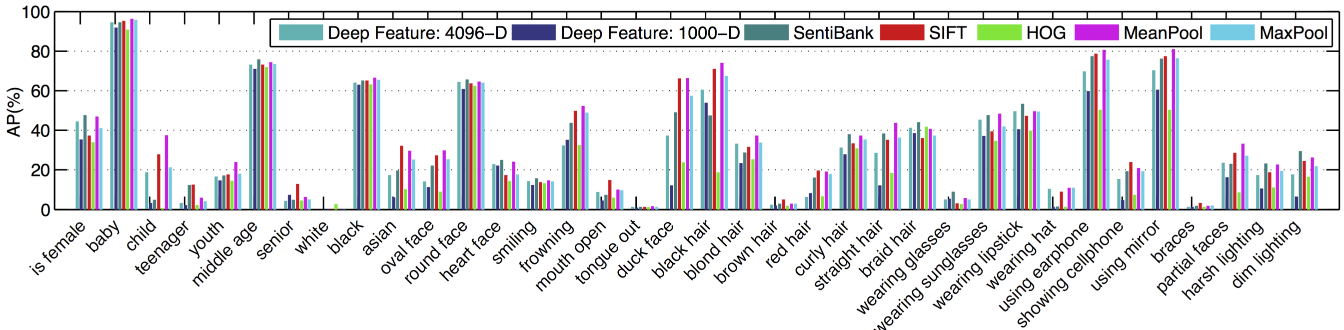

Figure 3: Attribute prediction performance of different features on the Selfie dataset.

influence the popularity of selfie images? We discuss the effect of attributes in section 6. Here we study object categories of ImageNet [6] and concepts associated with ANPs in SentiBank [2]. To evaluate the correlation of each object/concept with popularity of selfie images, we use 1000-D output of [7], pre-trained on ImageNet dataset, and 2089-D output of SentiBank ANP concept detectors as mid-level features. We then train an L_2-regularized support vector regression (SVR) to predict the popularity score of the images. Regression coefficients corresponding to different objects/concepts indicate their correlation to the popularity. We use object/concept detector responses with at least 0.5 confidence. This assures that we only consider images in which a particular object/concept is confidently detected. We observed that among ImageNet object categories: *maillot, lab coat, jersey, fur coat, brassiere, wig, abaya, hair spray, suit, sunglasses, and lipstick* (in decreasing order) are the most relevant objects to the popularity of selfies. Among different ANPs in SentiBank: *sexy dress, lovely dress, fancy dress, traditional tattoo, smiling baby, shiny hair, sexy girls, cute baby, strong legs, stupid hat and happy baby* (in decreasing order) are the most relevant concepts to the popularity score. Figure 4 illustrates the normalized regression coefficients (in decreasing order) obtained from training SVR.

Popularity Score Prediction: Observing the performance of different features in attribute prediction, we are also interested to see how they perform in predicting the popularity of selfie images. Thus, we randomly divided the entire Selfie dataset using 3-fold cross validation where 2 folds were used to train an L_2-regularized SVR and we tested on the third fold. We evaluate the quality of different features in terms of Spearman's rank correlation between the predicted and the actual popularity (generated by [4]). Using SIFT, we achieved 0.40 rank correlation. The 4096-D

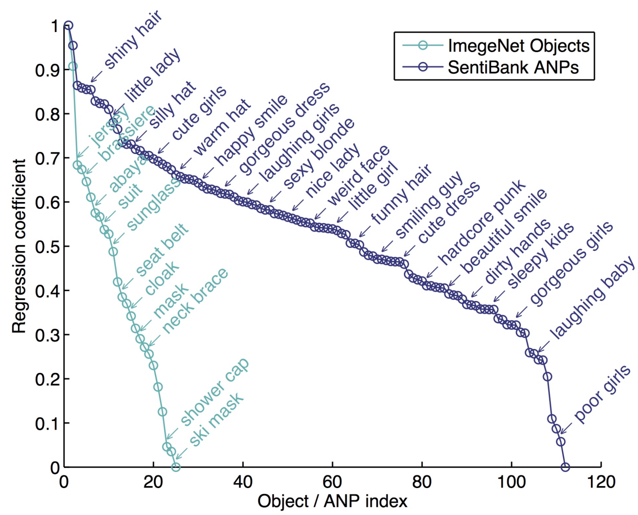

Figure 4: Normalized regression coefficients of SVR for popularity score prediction.

deep features and 2089-D SentiBank features resulted in 0.41 and 0.54 rank correlation, respectively. While max pooling of these features was not helpful (0.49 rank correlation), we observe that mean pooling can further boost the rank correlation to 0.55.

5. SENTIMENT-POPULARITY CORRELATION

In this section we explore the correlation between the popularity of selfie images and their implied sentiments. Each ANP in SentiBank is associated with a sentiment measure where negative, close to zero and positive numbers depict negative, neutral and positive sentiments, respectively. Out of 2,089 ANPs, we manually select 126 of them that are

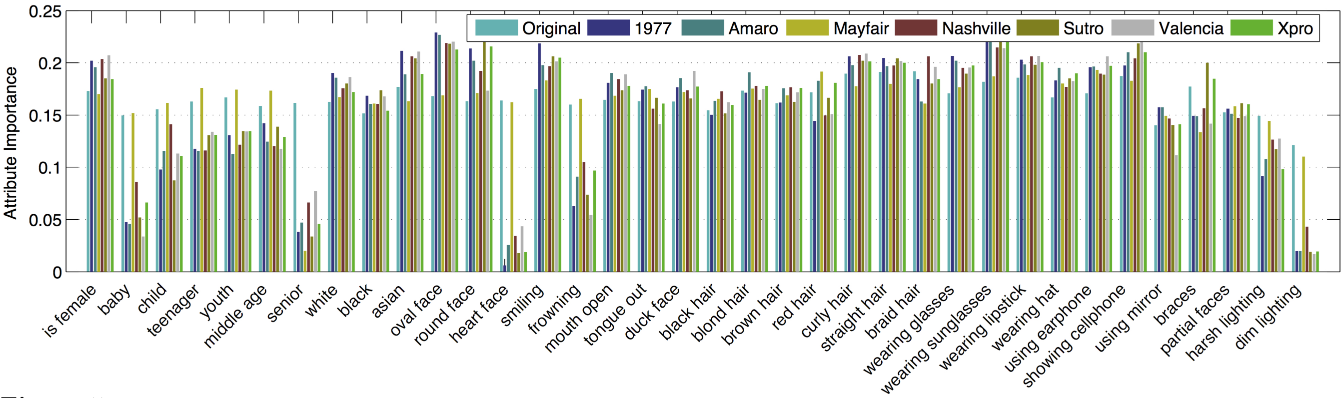

Figure 5: Importance of different attributes in predicting popularity, employing different Instagram filters. *Original* indicates no filter is applied.

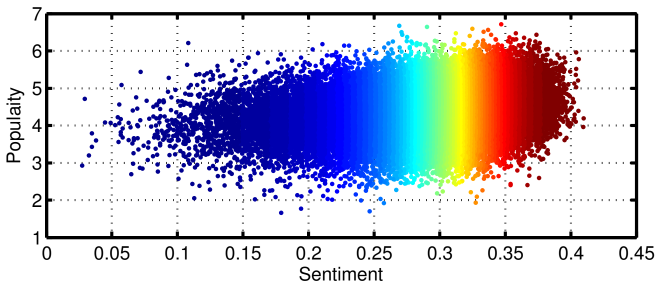

Figure 6: Sentiment-popularity scatter plot.

relevant to selfie images and their corresponding detectors have acceptable performance (AP$\geq$ 0.3). We compute implied sentiment of a selfie image I as $\mathcal{S}(I) = \sum_i \gamma_i s_i \omega_i^T x$, where γ_i, s_i and ω_i, respectively, represent the AP, sentiment measure and linear detector corresponding to the i^{th} ANP. Given x as the feature representation of image I, $\omega_i^T x$ is the confidence score for which i^{th} ANP concept appears in the image I. We generate the scatter plot of the popularity versus sentiment using our entire dataset. Figure 6 illustrates the scatter plot in which a more positive sentiment, on average, results in a higher popularity. We observe, on average, up to 65% higher popularity comparing the two ends of the sentiment spectrum, shown in color. Another interesting observation is how the range of popularity changes as the sentiment measure increases. From bluish to green/yellow parts of the spectrum, increasing sentiment yields a larger range of popularity score (the blue cone). However, moving toward more reddish parts of the spectrum, the range of the popularity decreases. Therefore, we conclude that while at two ends of the sentiment spectrum there exist a direct correlation between sentiment and popularity, this is not true for the middle of the spectrum. In other words, unless the sentiment is too high or too low, one cannot estimate the popularity based on the sentiment with high precision.

6. EFFECT OF POST-PROCESSING ON POPULARITY

This section addresses the last question that we proposed in the beginning of this paper: How does post-processing, such as applying different Instagram filters, influence the popularity of selfies? We randomly select about 10,000 images from our Selfie dataset and apply 7 different Instagram filters to them. We observe that for a given selfie, applying some filters boosts the popularity while others have a counter-effect. We found that there is no definite ranking of filters in terms of improving the popularity, rather ranking varies from one image to another. Figure 5 shows the relevance of different attributes to popularity of selfies when various Instagram filters are applied. Therefore, the choice of the most effective filter varies according to the content of a selfie. We obtain the importance of different attributes using SVR in a similar strategy discussed in section 4.

7. CONCLUSION

In this work, for the first time, we address the Selfie phenomenon from a machine learning and computer vision perspective. We conduct a series of experiments addressing four distinct questions about selfie images. Our work sheds light on how to take a *good selfie*, a selfie that becomes popular and delivers a positive sentiment. We collected, to the best of our knowledge, the first Selfie dataset for research purposes with more than 46,000 images, annotated with 36 relevant attributes. Finally, we believe that our work creates novel opportunities for researchers in other disciplines to extend their studies on the selfie phenomenon to a larger scale.

8. REFERENCES

[1] S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces engage us: Photos with faces attract more likes and comments on instagram. In *ACM CHI*, 2014.

[2] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *ACM MM*, 2013.

[3] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *CVPR*, pages 3304–3311, 2010.

[4] A. Khosla, A. Das Sarma, and R. Hamid. What makes an image popular? In *ACM WWW*, 2014.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge, 2014.

[7] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *ICLR*, 2014.