Accepted Manuscript

Predicting First Impressions with Deep Learning

Mel McCurrie, Fernando Beletti, Lucas Parzianello, Allen Westendorp, Samuel Anthony, Walter J. Scheirer

Please cite this article as: Mel McCurrie, Fernando Beletti, Lucas Parzianello, Allen Westendorp, Samuel Anthony, Walter J. Scheirer , Predicting First Impressions with Deep Learning. Imavis (2018), doi:10.1016/j.imavis.2018.06.010

Predicting First Impressions with Deep Learning

Mel McCurrie[1], Fernando Beletti[1], Lucas Parzianello[1], Allen Westendorp[1],
Samuel Anthony[2,3], and Walter J. Scheirer[1]

[1]Department of Computer Science and Engineering, University of Notre Dame
[2]Department of Psychology, Harvard University [3]Perceptive Automata, Inc.

Abstract—Describable visual facial attributes are now commonplace in human biometrics and affective computing, with existing algorithms even reaching a sufficient point of maturity for placement into commercial products. These algorithms model objective facets of facial appearance, such as hair and eye color, expression, and aspects of the geometry of the face. A natural extension, which has not been studied to any great extent thus far, is the ability to model subjective attributes that are assigned to a face based purely on visual judgments. For instance, with just a glance, our first impression of a face may lead us to believe that a person is smart, worthy of our trust, and perhaps even our admiration — regardless of the underlying truth behind such attributes. Psychologists believe that these judgments are based on a variety of factors such as emotional states, personality traits, and other physiognomic cues. But work in this direction leads to an interesting question: how do we create models for problems where there is only measurable behavior? In this paper, we introduce a convolutional neural network-based regression framework that allows us to train predictive models of crowd behavior for social attribute assignment. Over images from the AFLW face database, these models demonstrate strong correlations with human crowd ratings.

Index Terms—Psychophysics, Face Attributes, Convolutional Neural Networks

1. Introduction

In human attribute modeling there often exists a disparity between the way humans describe humans and the way computational models describe humans. A large amount of describable attribute research in computer vision concentrates on objective traits. For example, work using the popular CelebA dataset [22], [30], [44], [42] applies different methods to model traits such as "Male" and "Bearded" with binary annotations. Beyond objective attributes, it is possible to model more subjective traits such as expression [12], [7], attractiveness [17], and humorousness [21], but research often overlooks the important interrelation between attribute modeling and social psychology. Enabling computers to make accurate predictions about objective content and enabling computers to make human-like judgments about subjective content are both necessary steps in the development of machine intelligence. Here we focus on the latter.

Specifically, we concentrate on descriptions of the face, as an abundance of social psychology research demonstrates a human tendency to make judgments in social interactions based on the faces of fellow humans [33], [40], [1]. Popular human characteristics of academic interest closely related to these social interactions include emotion [24], attractiveness [1], trustworthiness [37], [40], [29], [8], dominance [33],
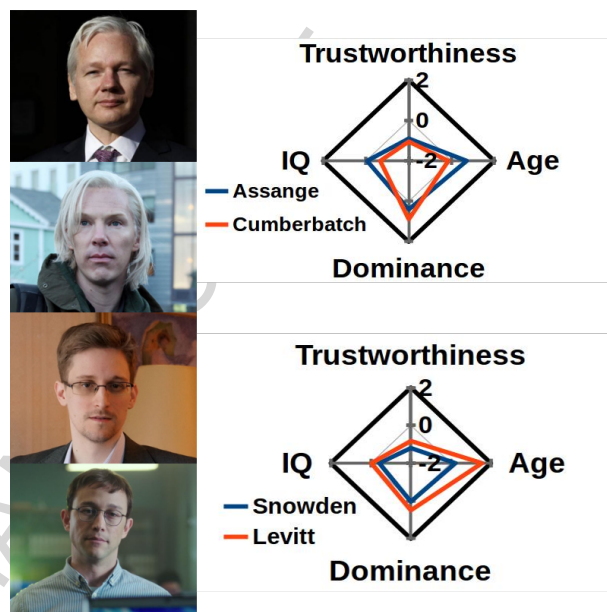

Fig. 1: **Graphical Abstract** Computational modeling of social attributes allows us to predict what the crowd might say about a face image. In this image we graphically compare the attribute predictions for Julian Assange and Benedict Cumberbatch, who plays Assange in the movie *The Fifth Estate*, as well as the predictions for Edward Snowden and Joseph Gordon-Levitt, who plays Snowden in the movie *Snowden*. Specifically looking at these images, our models output similar predictions between the subjects and their actors. The radar plots above reflect the output of a face processing pipeline, where faces are detected, aligned, and then processed through a deep convolutional neural network regressor that models a particular social attribute. This regression framework is the main contribution of our work. For this image we display the predictions' z-scores with respect to the training data.

[24], sociability, intelligence, and morality [1]. Psychologists often specifically concentrate on trustworthiness, dominance, and intelligence because they represent comprehensive abstract qualities that humans regard in each other. Alexander Todorov, one of the foremost psychologists studying these social judgments uses dominance and trustworthiness as the basis of many in-depth studies of human judgment [37], [38], [36]. Ultimately he finds that most other recogniz-