# Give Me One Portrait Image, I Will Tell You Your Emotion and Personality

Songyou Peng, Le Zhang, Stefan Winkler, Marianne Winslett
Advanced Digital Sciences Center, University of Illinois at Urbana-Champaign, Singapore
{songyou.peng, zhang.le, stefan.winkler}@adsc-create.edu.sg winslett@illinois.edu

## ABSTRACT

Personality and emotion are both central to affective computing. Existing works address them individually. In this demo we investigate if such high-level affect traits and their relationship can be jointly learned from face images in the wild. To this end, we introduce an end-to-end trainable and deep Siamese-like network. At inference time, our system can take one portrait photo as input and predict one's Big-Five apparent personality as well as emotion attributes. With such a system, we also demonstrate the feasibility of inferring the apparent personality directly fro emotion.

## CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Machine learning**;

## KEYWORDS

Personality; Emotion; Emotion-to-Personality; Deep Learning

**ACM Reference Format:**
Songyou Peng, Le Zhang, Stefan Winkler, Marianne Winslett. 2018. Give Me One Portrait Image, I Will Tell You Your Emotion and Personality. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3240508.3241384

## 1 INTRODUCTION

Proliferation of cameras, availability of cheap storage and rapid developments in high-performance computing have spurred applications in the field of Human-Computer Interaction (HCI), in which affective computing plays an inevitable role. For instance, in video-based interviews, automatically computed personalities of candidates can serve as an important cue to assess their qualifications. However, affective computing remains a challenging problem in both computer vision and psychology despite many years of research.

Here we study the fundamental problem of analyzing apparent personality, emotion, and their relationship. Apparent personality[1] reflects the coherent patterning of apparent behavior, cognition and desires (goals) over time and space, while emotion is an integration of feeling, action and wants at a particular time and location [6]. We can consider this emotion-to-personality relationship as weather to climate, i.e. what one expects is personality while what one observes in a particular moment is emotion. Although they have distinct definitions, their relationship has been revealed previously. Eysenck's personality model [2] showed that extraverts require more external stimulations than introverts; in other words, extraversion is accompanied by low cortical arousal. He also concluded that neurotics could be more sensitive to external stimulation and easily become upset or nervous due to minor stressors.

Deep CNNs reign undisputed as the new *de-facto* method for face based applications like face recognition [4, 7] or alignment [11]. This motivates us to study the following questions: how transferable are deeply learned face representations for emotion and personality analysis? And is it beneficial to explore emotion, personality and emotion-to-personality relationship in a single deep CNN? These tasks are non-trivial and the most significant challenges are:

(1) Lack of datasets which encompass both emotion and personality annotations for learning a rich representation for them and their relationship. In particular, existing datasets only have either emotion or personality attributes.
(2) Different datasets are collected under different scenarios which may exhibit significant variations in content, scale or background, which leads to divergent statistical distributions. Moreover, personality is usually labelled at the video level, whereas only one frame is needed for emotion annotation.

Consequently, encapsulating both levels of understanding into a single network triggers more challenges. We address them by proposing an end-to-end trainable and deep Siamese-like [1] network for the joint prediction of emotion (in terms of arousal and valence) and "Big 5" personality traits, which we describe in the next Section.

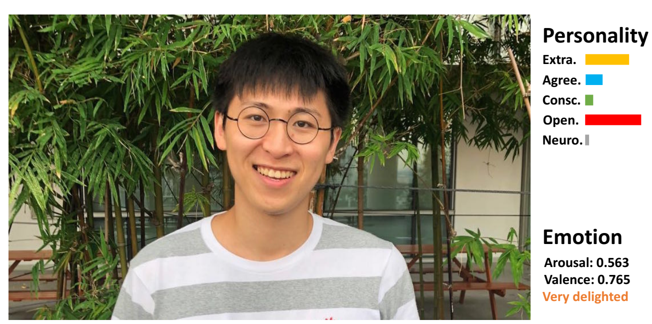


**Figure 2: Prospective demo interface**

[1]For the sake of simplicity, we use the term "personality" to represent apparent personality in the rest of the paper.


*MM '18, October 22–26, 2018, Seoul, Republic of Korea*

ACM ISBN 978-1-4503-5665-7/18/10.
https://doi.org/10.1145/3240508.3241384

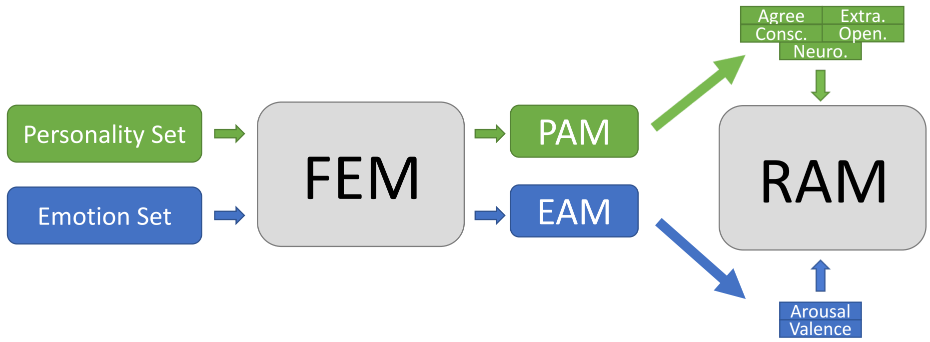


**Figure 1: Architecture of the proposed deep network.**

## 2 PROPOSED NETWORK

An overview of the proposed network is illustrated in Figure 1. We first use *MTCNN* [8] to detect and align faces for both personality [5] and emotion [10] datasets. For the videos in the personality dataset, a sparse sampling strategy is employed, where a video is split into 3 snippets and one single frame is randomly sampled from each of them. The personality network consists of a feature extraction module (*FEM*) and personality analysis module (*PAM*) to predict the Big-Five personality factors (Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness). *SphereFace* [3] is applied for *FEM*. An average operation is used to aggregate raw personality scores before feeding them into *PAM*. Similarly, the emotion network shares the *FEM* module with the personality network and has its own emotion analysis module (*EAM*) aimed at estimating arousal and valence.

In addition, two datasets exhibit vastly different statistical distributions; as a result, features extracted from each domain without pursuing coherence between them may present significant discrepancies. To tackle this problem, an adversarial-like loss function is employed to promote representation coherence between heterogeneous dataset sources, similar to [8].

Moreover, personality usually represents an integration of emotion over time and space [6]. [9] established the relationship between personality and affective attributes, but not in a computational setting. In our system, an emotion-to-personality relationship analysis module (*RAM*) is employed to predict the 5 personality traits from 2 emotion attributes. With *RAM*, our network provides a promising solution for automatically annotating frame-level personality directly from people's emotions.

Every module of the network is differentiable, allowing for end-to-end optimization of the whole system. The learning process of the proposed network aims to minimize the following overall loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{per} + \lambda_2 \mathcal{L}_{emo} + \lambda_3 \mathcal{L}_{coh} + \lambda_4 \mathcal{L}_{RAM}$$

$\mathcal{L}_{per}$ and $\mathcal{L}_{emo}$ represent the prediction errors of *PAM* and *EAM* on the personality and emotion datasets separately, while $\mathcal{L}_{coh}$ stands for the loss of the representation coherence between the two datasets, and $\mathcal{L}_{RAM}$ the loss of the prediction from emotion to personality.

## 3 CONCLUSION

For the first time, the feasibility of jointly analyzing the apparent personality, emotion and the emotion-to-personality relationship within a single deep neural network is investigated. This is challenging due to the lack of datasets which encompass both emotion and personality annotations. The proposed joint training of both emotion and personality networks brings about more generalizable representations for both tasks. We also find that one's personality can be well predicted by emotions.

As for the demo, our system is capable of taking a single photo for personality and emotion prediction, as well as providing frame-level annotations of a video clip or the live stream from a webcam.

## ACKNOWLEDGMENTS

This research is supported by the SERC Strategic Fund from the Science and Engineering Research Council (SERC), A*STAR (project no. a1718g0046).

## REFERENCES

[1] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. 1994. Signature verification using a "Siamese" time delay neural network. In *Proc. NIPS*.
[2] Hans J Eysenck. 1950. *Dimensions of Personality*. Vol. 5. Transaction Publishers.
[3] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. 2017. Sphereface: Deep hypersphere embedding for face recognition. In *Proc. CVPR*.
[4] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and others. 2015. Deep Face Recognition. In *Proc. BMVC*.
[5] Víctor Ponce-López, Baiyu Chen, Marc Oliu, Ciprian Corneanu, Albert Clapés, Isabelle Guyon, Xavier Baró, Hugo Jair Escalante, and Sergio Escalera. 2016. Chalearn LAP 2016: First round challenge on first impressions – Dataset and results. In *Proc. ECCV Workshop*.
[6] William Revelle and Klaus R Scherer. 2009. Personality and emotion. *Oxford Companion to Emotion and the Affective Sciences* (2009), 304–306.
[7] Yi Sun, Yuheng Chen, Xiaogang Wang, and Xiaoou Tang. 2014. Deep learning face representation by joint identification-verification. In *Proc. NIPS*.
[8] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. 2015. Simultaneous deep transfer across domains and tasks. In *Proc. ICCV*.
[9] Kathy A Winter and Nicholas A Kuiper. 1997. Individual differences in the experience of emotions. *Clinical Psychology Review* 17, 7 (1997), 791–821.
[10] Stefanos Zafeiriou, Dimitrios Kollias, Mihalis A Nicolaou, Athanasios Papaioannou, Guoying Zhao, and Irene Kotsia. 2017. Aff-wild: Valence and arousal in-the-wild challenge. In *Proc. CVPR Workshop*.
[11] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.