Visualizing Apparent Personality Analysis with Deep Residual Networks

Yağmur Güçlütürk, Umut Güçlü, M. Pérez, H. J. Escalante, X. Baró, I. Guyon, C. Andujar,
J. Jacques Jr, M. Madadi and S. Escalera, M. A. J. van Gerven, R. van Lier
Radboud University, Nijmegen, the Netherlands, University of Barcelona, CVC, UOC, Barcelona, Spain,
INAOE, Mexico, UPSud/INRIA UP-Saclay, ChaLearn, Paris, France, UPC, BarcelonaTech

ICCV17
International Conference on Computer Vision 2017

Explaining first impressions analysis

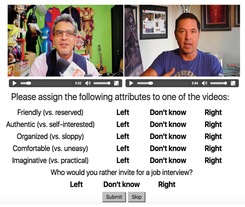
One of the main goals of the studies of computer vision-based apparent personality trait recognition and analysis is to increase our understanding of the underlying psychological phenomena through modeling.

Mechanisms for interpreting, and explaining model predictions are highly demanded in this area.

We devise and evaluate the performance of a state of the art methodology for apparent personality estimation.

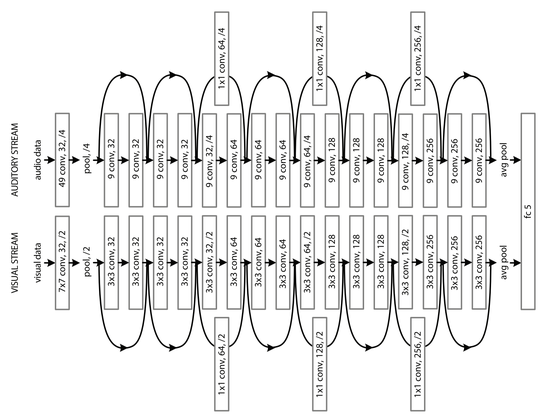
First impressions data set

A novel data set comprising 10,000 clips, 15s each, labeled with big-5 apparent personality traits was considered in the study.



Apparent personality estimation model

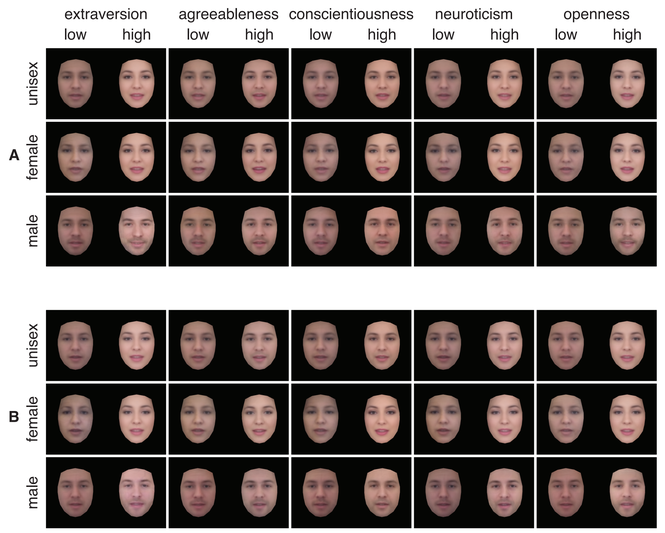
We consider a two-stream (audio-visual) deep residual network for our analysis. This model obtained outstanding performance (above 90% of accuracy) in the first impressions challenge.



We applied different visualization analyses to interpret and understand the functioning of the model.

Representative faces of apparent traits

We created representative images of faces for the highest and lowest levels of each trait based on **(A)** the annotations of the test set videos as well as **(B)** the predictions by the model.

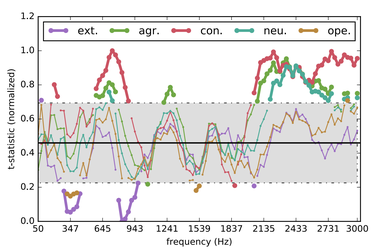


Visual inspection of the results revealed:

- Average female faces with high levels of all traits seemed to be more colorful with higher contrast compared to those with low levels
- Regarding unisex avg. faces, a bias for female faces for the high levels and male faces for the low levels was observed

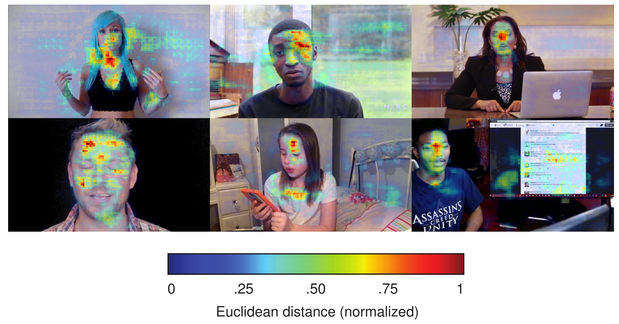
Occlusion analysis

We systematically masked the visual or audio inputs to the network and measured the changes in predictions as a function of location, predefined region or frequency band.



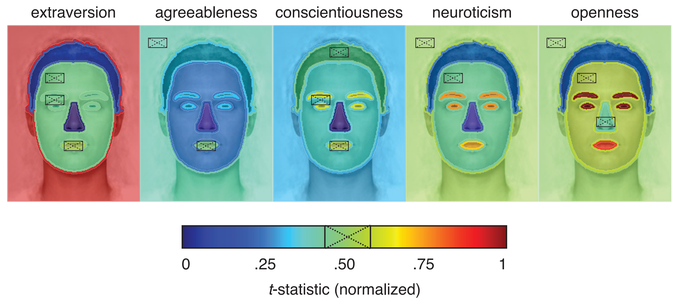
Audio. Each line shows the changes in the prediction of the corresponding trait as a function of frequency resulting from systematically masking the videos. Grayed out region indicates the changes that are not significantly above zero.

Pixel level occlusion analysis. We systematically masked the input with 10x10 pixel square masks that were centered on every fourth point in the spatial axes.



- Faces in the video frames drove the predictions of the network the most
- Objects in the background were also observed to have an influence, but to a lesser extent.

Segment level occlusion analysis. Masks correspond to specific regions of the face (+background)



- Each region modulated at least one trait, and different traits were modulated by different regions.

Trait recognition system

We designed a web application to recognize apparent personality traits (Big Five), which was presented as a demonstration at NIPS 2016.