Visualizing Apparent Personality Analysis with Deep Residual Networks

Yağmur Güçlütürk, Umut Güçlü, Marc Perez, Hugo Jair Escalante Balderas, Xavier Baró, Isabelle Guyon, Carlos Andujar, Julio Jacques Junior, Meysam Madadi, Sergio Escalera, et al.

Visualizing Apparent Personality Analysis with Deep Residual Networks

Yağmur Güçlütürk, Umut Güçlü
Radboud University,
Nijmegen, the Netherlands
{y.gucluturk,u.guclu}@donders.ru.nl

Marc Pérez
University of Barcelona
Barcelona, Spain
marcperez1993@gmail.com

Hugo Jair Escalante
INAOE, ChaLearn
Puebla, Mexico
hugojair@inaoep.mx

Xavier Baró
UOC, CVC
Barcelona, Spain
xbaro@uoc.edu

Isabelle Guyon,
UPSud/INRIA UP-Saclay, ChaLearn
Paris, France
guyon@clopinet.com

Carlos Andujar
UPC, BarcelonaTech
Barcelona, Spain
andujar@cs.upc.edu

Julio Jacques Junior, Meysam Madadi and Sergio Escalera
University of Barcelona, CVC
Barcelona, Spain
{juliojj, meysam.madadi, sergio.escalera.guerrero}@gmail.com

Marcel A. J. van Gerven, Rob van Lier
Radboud University,
Nijmegen, the Netherlands
{m.vangerven,r.vanlier}@donders.ru.nl

Abstract

Automatic prediction of personality traits is a subjective task that has recently received much attention. Specifically, automatic apparent personality trait prediction from multimodal data has emerged as a hot topic within the filed of computer vision and, more particularly, the so called "looking at people" sub-field. Considering "apparent" personality traits as opposed to real ones considerably reduces the subjectivity of the task. The real world applications are encountered in a wide range of domains, including entertainment, health, human computer interaction, recruitment and security. Predictive models of personality traits are useful for individuals in many scenarios (e.g., preparing for job interviews, preparing for public speaking). However, these predictions in and of themselves might be deemed to be untrustworthy without human understandable supportive evidence. Through a series of experiments on a recently released benchmark dataset for automatic apparent personality trait prediction, this paper characterizes the audio and visual information that is used by a state-of-the-art model while making its predictions, so as to provide such supportive evidence by explaining predictions made. Additionally, the paper describes a new web application, which gives feedback on apparent personality traits of its users by combining model predictions with their explanations.

1. Introduction

Recent studies in the literature show that first impressions about personalities of people based on analyses of their faces tend to change from one photo to another (i.e., the ratings vary with respect to context) [16], making the evaluation a very complex process. Similarly, Sutherland et al. [13] demonstrated that the emotional expression of the face and the viewpoint of the photograph were important factors influencing the within-person variability of the impressions regarding the trustworthiness, dominance, and attractiveness traits.

According to Todorov et al. [17], independent of their accuracy, trait inferences affect important social outcomes. For example, positive traits such as intelligence are more likely to be attributed to more attractive individuals [14]. However, contrasting observations have also been made, such that more attractive scientists are perceived to be less competent than their less attractive peers [4]. Such con-

tradictory results regarding the influence of different factors on first impression formation highlight the need for assumption-free exploratory investigations of the underlying factors.

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. Previous work found in the literature proposed to explore the relationship between image features and first impressions [20, 18] in order to determine which region is more relevant to predict different personality traits. However, facial features (e.g., region based or action units analysis) or facial expressions [15] have been more in the focus of such analysis, and the effects of other features, such as the background region and/or audio have not been as adequately studied. In this work, we analyze the functioning of a state-of-the-art method for apparent personality trait recognition, which utilizes deep residual networks and audiovisual information. Specifically, we aim to reveal insightful information that the model exploits for recognizing personality traits. We believe this information can be used to explain the recommendations made by the considered model, making it more interpretable and, henceforth, more reliable for decision support systems based on apparent personality analysis.

The remainder of this paper is organized as follows. The next section elaborates on the considered data set and the associated first impressions challenge. Section 3 presents the underlying method for apparent personality traits estimation. Section 4 provides an analysis on the visualization of information associated to personality trait estimation. Section 5 presents the web based application for personality analysis. Finally, Section 6 outlines conclusions derived from this work.

2. The First Impressions Data Set

For our study we consider a recently released data set for personality analysis. The so called, *first impressions* data set was released in the context of a series of academic competitions [3, 10, 2] focused on the analysis of personality and adequacy of individuals for being invited to a job interview.

The first impressions[1] data set is made up of 10,000 video clips with an average duration of 15s. These clips were extracted from more than 3,000 different YouTube videos of English speaking people, and most of them were recorded in a video blog format. People in videos have different gender, age, nationality, and ethnicity. Furthermore, the backgrounds of the videos are not uniform, making the task of inferring apparent personality traits even more challenging. Figure 1 shows snapshots of sample videos from the data set.

[1]Data set available at http://chalearnlap.cvc.uab.es/

Figure 1. Snapshots of sample videos from the First Impressions data set [10].

Videos have personality traits and a so called "job-interview variable" labels. Amazon Mechanical Turk (AMT) was used for collecting the data to generate these labels. In order to guarantee the reliability of the labels, all rankings provided by the AMT workers were converted into normalized real valued scores using a principled procedure (see [1] for details). Personality traits that were used in the dataset were those from the Five Factor Model (also known as the *Big Five*). The Five Factor Model is a dominant paradigm in personality research, and models human personality along five dimensions: *Extraversion, Agreeableness, Conscientiousness, Neuroticism* and *Openness*. AMT workers labeled each video based on how they perceived the individual in the video with respect to these five personality traits as well as an additional variable indicating whether the person should or should not be invited to a job interview (the "job-interview variable"). For further information regarding the data set, the reader is referred to [10] where the data set is described in more details.

The first impressions data set was formerly used in the context of an academic challenge [10, 3] that aimed to develop methods for predicting the big 5 personality traits. More recently, an extended version of the data set was used in a competition that had as goal to develop methods for predicting whether a subject should or should not be invited to an interview. Additionally, participants had to explain the recommendations of their models [2].

3. Deep Residual Networks for Multimodal Apparent Personality Trait Recognition

Here, we consider a model described in [6, 7]. This model was ranked third in the ChaLearn First Impressions Challenge 2016 [12]. Please note that the difference in performance with the top ranked methodology was negligibly small: [6] obtained a test-set accuracy of 0.9109, whereas the winner [22] obtained an accuracy of 0.9129. Making this model very competitive and representative of the state-of–the-art. The rest of this section briefly describes such methodology. We refer the reader to [6, 7] for a detailed description.

This model made use of a deep learning methodology, which took advantage of both visual and auditory information available in the data set. Specifically a two stream deep residual network [8] was developed. The network was trained end-to-end without the usage of any feature engi-

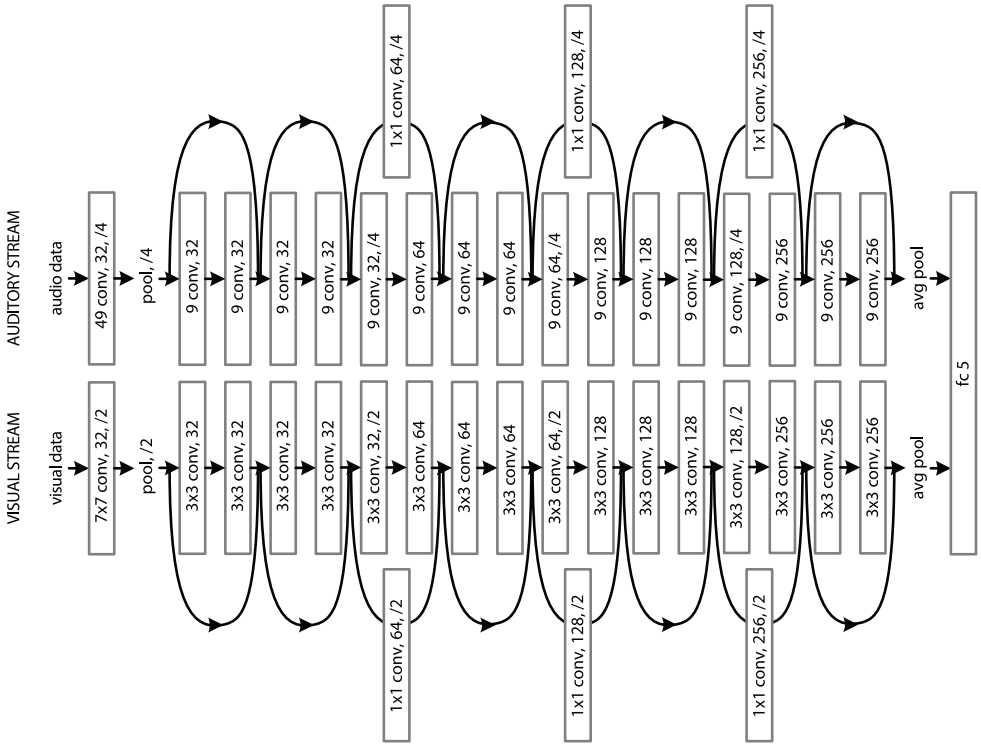

Figure 2. Illustration of the network architecture. Each box denotes a convolutional neural network layer in the following notation: 'kernel size (height × width for visual data and width for auditory data)' conv, 'kernel number' / 'stride'. Arrow connecting non consecutive layers denote residual connections. Figure adapted from [6, 7].

neering, both for the audio and the visual components of the input videos. This characteristic of the network is ideal for analysis and interpretation of the features that contributed to the predictions of the network as no assumptions were made and no information is discarded in advance.

The network is made up of an auditory stream of a 17 layer deep residual network, a visual stream of another 17 layer deep residual network and audiovisual fully-connected layer on top. The architecture is depicted in Figure 2. The network was trained by minimizing the mean absolute error loss function between ground truths and predictions.

4. Visualizing traits

The next section describes the set of experiments that were conducted with this model with the aim of revealing informative patterns from data that can explain the predictions of the model that was outlined in the previous section. First, we perform an analysis on prototypical faces associated to different traits. Next, occlusion analysis is performed to detect regions of interest in images, aiming to explain trait predictions. Finally, a similar analysis is performed on the audio modality.

4.1. Representative Faces of Different Traits

We created representative images of faces for the highest and lowest levels of each trait based on the annotations of the test set videos as well as their predictions by the model. As a result, typical faces having different levels of different traits in the eyes of the annotators as well as the model were visualized. To create such an image for the highest level of a particular trait based on the annotations, the following procedure was used:

First, 100 videos were selected such that their corresponding annotations for the trait was the highest. Then, the first and last frames of these videos were extracted. Next, the faces in these frames were linearly aligned and triangulated (Delaunay triangulation) based on an automatically estimated set of 77 facial key points, and a template was created by averaging the coordinates of each triangle over the faces. Finally, the triangles in each face were separately aligned to the corresponding triangles in the template and averaged over the faces. This procedure was performed not only for all faces in the selected videos altogether, but also for the female and male faces in the selected videos separately.

Images for the lowest levels of a trait and/or based on

the predictions were created analogously. For example, 100 videos were selected such that their corresponding predictions for the trait were the lowest.

Figure 3 shows the results of this procedure. A common feature of all average face groups (i.e. female, male and unisex, as well as those based on the annotations and predictions) was the existence of prominent differences between the average faces having low and high levels of each trait. Similar to the results of previous attempts of visualizing the facial features contributing to impressions of the Big Five personality traits [19, 9], differences between high and low levels of all traits were observed to be prominent in terms of the facial structure, image properties and expressions of the average faces.

For instance, high levels of all traits were more bright and colorful, and with more positive expressions. This result is in line with [19], where the authors found that manipulations in shape space explained 82-90% of the variability of personality trait judgments and manipulations in color space explained 73-84% of the variability of personality trait judgments.

Visual inspection of the results revealed that the average female faces with high levels of all traits seemed to be more colorful with higher contrast compared to those with low levels, who were more uniformly colored with lower contrast. Furthermore, looking at the unisex average faces, a bias for female faces for the high levels and male faces for the low levels was observed, more so in the case of average faces based on annotations. However, the differences between the average faces having low levels of different traits were much more subtle especially for the female average faces. This was also the case for the average faces having high levels of different traits.

4.2. Explaining the Predictions Using Occlusion Analysis

In order to explore what influenced the predictions of the considered deep network, further experiments were performed. These experiments involved systematically masking the visual or audio inputs to the network and measuring the changes in predictions as a function of location, predefined region or frequency band [21]. The motivation behind these analyses was that if a certain location was driving the predictions, then masking it would either increase or decrease these predictions, enabling us to visualize the regions or audio frequencies that had the most effect for the predictions of each personality trait. Two analyses were performed, one considering raw pixels and another one focusing on segments. These are described in the rest of this subsection.

4.2.1 Experiment 1: Visual pixel-level occlusion analysis

Pixel-level occlusion analysis was performed by systematically masking the input with 10×10 pixel square masks that were centered on every fourth point in the spatial axes. First, for each video, the model was used to predict the five traits using only the first frame of its video track but its entire audio track.

For each mask location, the analysis was then performed as follows:

1. The video frame was covered with a 10×10 pixel gray square mask.
2. All traits were predicted using the masked frame while keeping the audio track fixed.
3. The change between the initial predictions and the masked predictions were calculated in terms of the Euclidean distance between the five trait predictions before and after masking.

The obtained changes were then visualized as a contribution map superimposed on the corresponding video frame. Visual inspection of the results revealed that the faces in the video frames drove the predictions of the network the most (Figure 4). The specific face regions that had the most influence varied substantially for each video frame. Furthermore, in some videos objects in the background were also observed to have an influence, but to a lesser extent.

4.2.2 Experiment 2: Visual segment-level occlusion analysis

The pixel-level occlusion analysis provided us with a general but rather coarse understanding of where in a video frame the most important information was located. Therefore, to further explain the results, we performed one more visual occlusion analysis experiment. This time rather than using square shaped masks at arbitrary locations of the video frame, masks that corresponded to semantically meaningful regions in the video frames were used. Furthermore, by performing the analysis per each trait, the question whether the predictions of different traits were driven by different factors was investigated.

In order to perform the segment-level occlusion analysis, video frames were segmented into the following six regions with a deep neural network: background region, hair region, skin region, eye region, nose region and mouth region. Skin region was defined as entire face without the other regions as well as the ears and the neck. Segmentation was performed using the method presented in [5]. The eye region was defined as the two eyes and the two eye brows. The mouth region was defined as the upper lip, inner mouth and

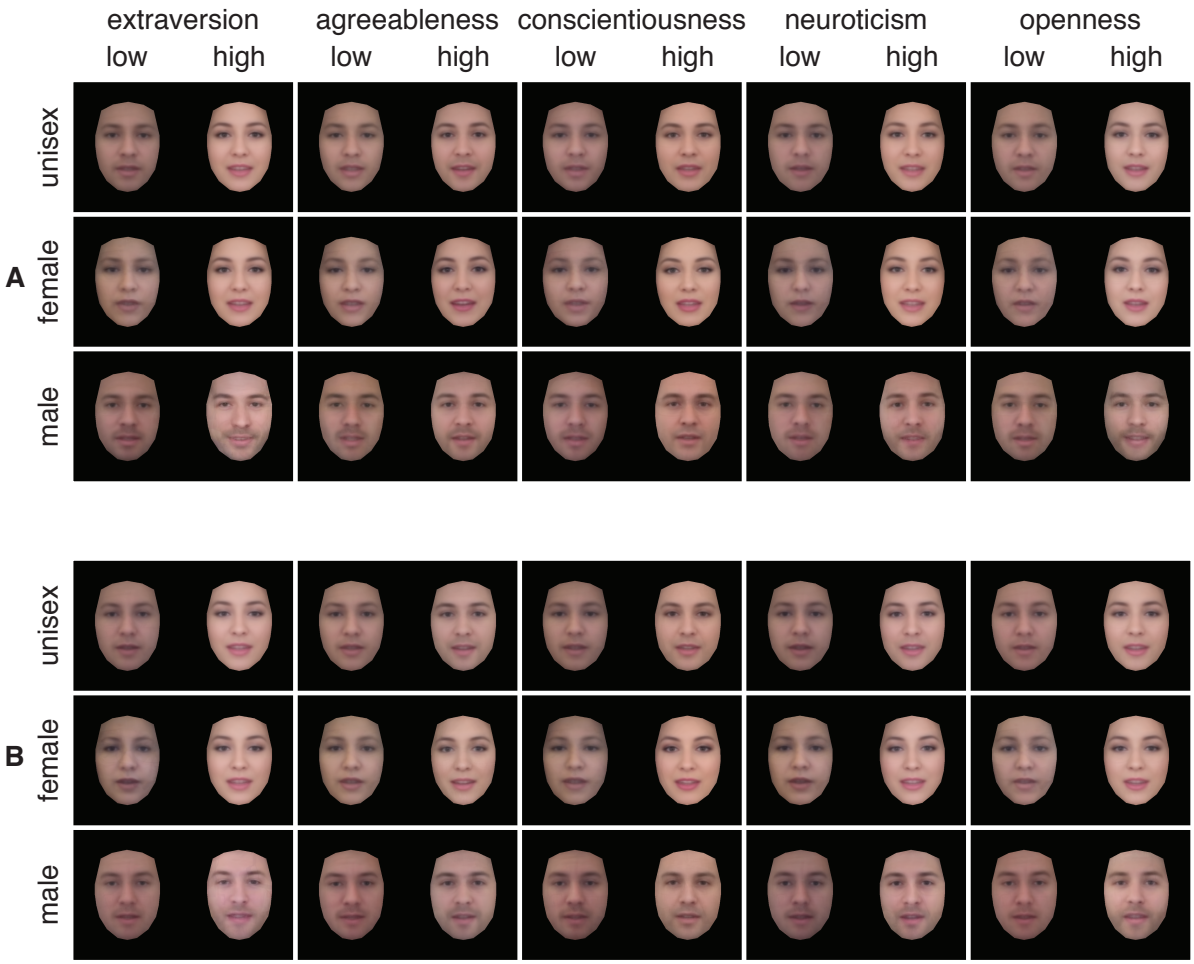

Figure 3. Representative faces of different traits. (A) Images created based on the annotations. (B) Images created based on the predictions.

Figure 4. Visual pixel-level occlusion analysis. Each image shows the changes in trait predictions as a function of location resulting from systematically masking a representative example input overlaid on the input itself. Masks are defined as 10×10 pixels centered on every fourth point in the spatial axes. Change is defined as the Euclidean distance between the predictions before and after masking the inputs.

lower lip. The background region was defined as all the remaining regions in a video frame which often included the other body parts such as the neck, hands, arms, etc., as well as the actual background of the video.

First, the model was used to predict the five traits from all videos in the test set. For each video, only the first frame of its video track but its entire audio track was used.

Then, the following operations were performed on the same data as above for each trait and region pair:

1. The region was covered with a gray mask.

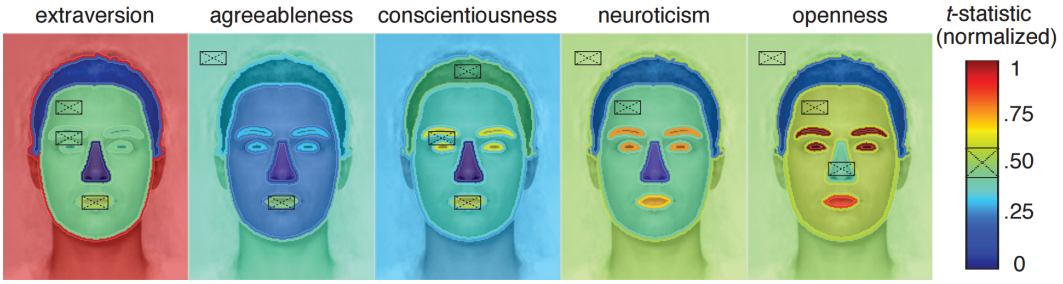

Figure 5. Segment-level occlusion analysis. Each image shows the changes in the prediction of the corresponding trait as a function of a predefined region resulting from systematically masking the videos overlaid on an average face (cold colors indicate the region of image that is less relevant do predict such trait). Masks are estimated with a separate deep neural network trained for segmenting faces to six regions. Change is defined as the effect size of the difference between the predictions before and after masking the videos. Crosses indicate the changes that are not significantly above zero.

2. The trait was predicted using the masked frame while keeping the audio track fixed.

3. The difference between the initial prediction and the masked prediction was taken.

4. The significance of the difference was assessed at 0.05 level with a Student's t-test over the videos. Bonferroni correction was used to control for multiple comparisons.

This procedure resulted in a t-value for each trait and region pair. It is important to note that the t-values do not indicate what specific features of a region influence the prediction of a trait. Rather they indicate how much the region influences the prediction of the trait on average. For brevity, we refer to a region significantly influencing the prediction of a trait as that region modulating that trait or that trait being modulated by that region. Figure 5 illustrates these t-values as contribution maps superimposed on an average face.

The most prominent pattern observed in these contribution maps is that each region modulated at least one trait, and different traits were modulated by different regions. Background, skin and mouth regions modulated the fewest traits. Occlusion of background region increased extraversion trait but decreased conscientiousness trait. Occlusion of skin region decreased agreeableness and conscientiousness traits. Occlusion of mouth region increased neuroticism and openness traits. In contrast, hair and nose regions negatively modulated all but one trait. The only traits that did not change by occlusion of these regions were conscientiousness and openness traits, respectively. Finally, eye region negatively modulated agreeableness trait, and positively modulated neuroticism and openness traits.

Although it is difficult to interpret these findings, our results correspond well with recent works on personality traits in the social psychology literature in general. For instance, similar to [11] we found that the overall appearance of a person affects the personality judgements regarding them. Both the face region and the background region which includes clothing, posture and body parts have been show to contain information regarding the apparent personality traits in both studies. Although their results were limited mostly to conscientiousness and extraversion traits, [9] showed that for different personality traits, different facial features were utilized in judgements regarding the apparent personality of unknown individuals. This is indeed comparable to the results of our segment-level occlusion analysis as illustrated in Figure 5.

4.2.3 Experiment 3: Audio frequency occlusion analysis

In the auditory occlusion analysis, the same procedure as the visual occlusion analysis was used except for keeping the frames fixed and covering the auditory track with a frequency mask. Briefly, t-values that indicate the effect of different frequency bands on the predictions of the traits were estimated. Frequency masks were defined as second order bandstop Butterworth filters with different center frequencies and bandwidths. Concretely, we used 100 masks, each with a center frequency linearly spaced between 50 Hz and 3000 Hz, and a bandwidth of 30 Hz.

Figure 6 illustrates the results of this analysis. Agreeableness, conscientiousness and neuroticism traits were influenced by the masking of a large number of frequency bands, while the extraversion and openness traits were affected by the masks to a lesser extent. Furthermore, results suggest that the traits agreeableness, conscientiousness and neuroticism were affected positively by audio signals containing both high and low frequencies, whereas extraversion and openness were only affected negatively and only by lower frequencies. Lower frequencies of human speech is known to correspond mostly to vowel sounds and higher frequencies above 1500 Hz correspond mostly to consonants. Relating these auditory occlusion analysis results to

human speech, it appears that our model's predictions about the extraversion and openness traits were mostly influenced by how the speakers in the videos uttered the vowels rather than the consonants, whereas in the case of the remaining personality traits, i.e. agreeableness, conscientiousness and neuroticism, the utterances of both the vowels and the consonants played a role. Furthermore, these results suggest that especially for the extraversion and openness traits, visual cues rather than the auditory cues might have been more dominant in the predictions of the model.

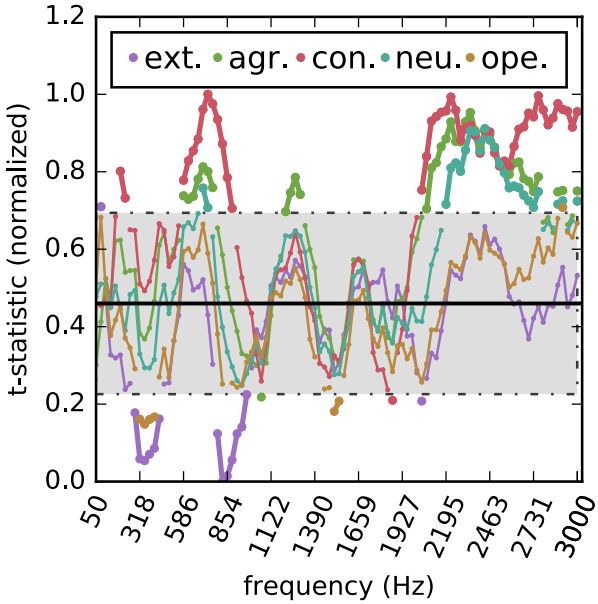

Figure 6. Audio frequency occlusion analysis. Each line shows the changes in the prediction of the corresponding trait as a function of frequency resulting from systematically masking the videos. Masks are defined as the 30 Hz bands centered around the frequencies. Change is defined as the effect size of the difference between the predictions before and after masking the videos. Grayed out region indicates the changes that are not significantly above zero. A result e.g. above the grayed out region shows that masking a frequency band resulted in a significant decrease in the value of the trait, suggesting that audio signals in those frequencies normally increase the value of the trait. Black solid line indicates the normalized t-statistic corresponding to a change of zero.

5. Trait recognition online system

Based on the first impressions data set for apparent personality trait recognition and analysis and the deep residual neural network explained in Section 3, we designed a web application to recognize apparent personality traits (Big Five), which was presented as a demonstration at NIPS 2016.

It is important to emphasize that the predictions of our system do not necessarily reflect the true personality traits of the individual, but the *apparent* personality traits. Also note that even trained humans find it difficult to determine the personality traits of others. The system rather provides an opportunity for individuals to learn about what other people would think of them after a very brief interaction. An example where such a tool can be used is as a preparation aid for a job interview (and as a system for supporting the decision of recruiters in the other way around). For example, the interviewees can record a short video of themselves talking about why they are good candidates for the job and obtain the predictions using our system (that is precisely the scenario considered in the job candidate screening challenge [2]). In case that they appear to be e.g. less extroverted than their actual personality, they can pay more attention to reflect this dimension of their personality more accurately during the actual job interview.

Figure 7. Interaction in our online trait recognition system is driven by an avatar representing the virtual interviewer.

Our web-based application consists of two parts. In the first part, after the presentation of instructions by an avatar (Figure 7), a short video from the user is recorded. This video should be in a similar format to a blog post and will consist of the volunteer freely talking about an issue. In the second part, this video will be automatically processed by the deep residual network. This processing runs on a remote server and takes approximately three seconds for the whole 15 s clip. The output consists of a rating from 0 to 100 for each one of the Big Five personality traits.

The web interface shows these ratings along with the adjustment of the traits to five job profiles. Furthermore, a list of Computer Vision and Machine Learning Researchers, with similar apparent personality to the one estimated for the user are shown. Figure 8 shows the output of the web application for a sample video. The user has also the option to request a summary personality report by email, which also includes the visualization procedure explained in this paper for the videos of the user. The client-side of the system is implemented in HTML5. Requirements for its usage include a computer with a webcam, microphone and internet connection. All heavy computational processing of the deep impression network runs on a GPU-based remote server.

This application was presented as part of the demonstration program at the NIPS conference, December 2016,

in Barcelona. More than 100 people participated in this demonstration. Table 1 shows a summary statistics of the predicted traits of the participants. Additionally, the participants were asked to fill-in a short survey for self-assessment of their personalities. Together with the survey results, the videos of the majority of the participants (those who gave consent) were stored for future analyses.

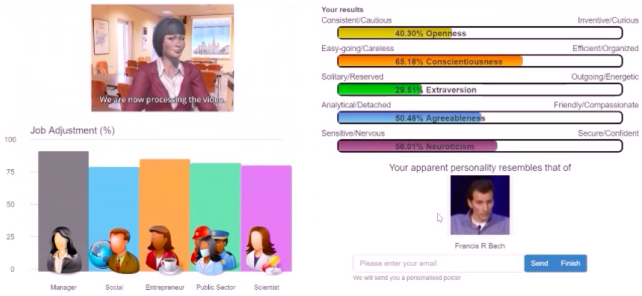

Figure 8. Snapshot of the output of the web interface for personality traits analysis.

Table 1. Summary statistics of the apparent personality trait predictions for the participants of the NIPS 2016 demo.

	min	max	mean $\pm$ std
agreeableness	0.3938	0.6241	0.5350 ± 0.0545
conscientiousness	0.3283	0.5630	0.4614 ± 0.0538
extraversion	0.2581	0.5782	0.4138 ± 0.0590
neuroticism	0.3696	0.6554	0.5262 ± 0.0561
openness	0.3338	0.6288	0.4786 ± 0.0608

6. Conclusions

Automated tools that can predict apparent personality traits from audio and/or visual data have a large potential for real world applications. Although such tools are limited in the sense that they can only predict apparent and not real personality traits, they can still be of great help in a number of scenarios.

This paper aimed to analyze what information apparent personality trait recognition models rely on when making thier predictions. To this end, we conducted a number of experiments to explain the predictions of such a state of the art model by characterizing the audio and visual information that drive these predictions. Furthermore, we described a new web application, which provides anyone the opportunity to receive feedback on their apparent personality traits.

Considering that explainability has recently emerged as a hot topic in machine learning, we foresee that our effort will motivate the investigation of the explanatory mechanisms for similar models in the field.

Acknowledgments

Marcel van Gerven was supported by a VIDI grant (639.072.513) from the Netherlands Organization for Scientific Research and a GPU grant (GeForce Titan X) from the Nvidia Corporation. Hugo Jair Escalante was supported by CONACyT under grants CB2014-241306 and PN-215546. This work has also been partially supported by the Spanish projects TIN2015-66951-C2-2-R and TIN2016-74946-P (MINECO/FEDER, UE), by the European Comission Horizon 2020 granted project SEE.4C under call H2020-ICT-2015, by the CERCA Programme/Generalitat de Catalunya, and NVIDIA GPU Grant Program.

References

[1] B. Chen, S. Escalera, I. Guyon, V. Ponce-Lopez, N. Shah, and M. O. Simon. Overcoming calibration problems in pattern labeling with pairwise ratings: Application to personality traits. In *ChaLearn Looking at People Workshop on Apparent Personality Analysis. ECCV Workshop proceedings, LNCS, Springer, 2016, in press.*, 2016. 2

[2] H. J. Escalante, I. Guyon, S. Escalera, J. J. Jr., M. Medadi, X. Baro, S. Ayache, E. Viegas, Y. Gcluturk, U. Guclu, M. van Gerven, and R. van Lier. Design of an explainable machine learning challenge for video interviews. In *Proc. of IJCNN 2017*, 2017. 2, 7

[3] H. J. Escalante, V. Ponce, J. Wan., M. Riegler, C. B., A. Clapes, S. Escalera, I. Guyon, X. Baro, P. Halvorsen, H. Müller, and M. Larson. Chalearn joint contest on multimedia challenges beyond visual analysis: An overview. In *ICPRW*, 2016. 2

[4] A. I. Gheorghiu, M. J. Callan, and W. J. Skylark. Facial appearance affects science communication. *Proceedings of the National Academy of Sciences*, 114(23):5970–5975, may 2017. 1

[5] U. Güçlü, Y. Güçlütürk, M. Madadi, S. Escalera, X. Baró, J. Gonzàlez, R. van Lier, and M. A. J. van Gerven. End-to-end semantic face segmentation with conditional random fields as convolutional, recurrent and adversarial networks. *CoRR*, abs/1703.03305, 2017. 4

[6] Y. Güçlütürk, U. Güçlü, M. A. van Gerven, and R. van Lier. Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition. *ECCV ChaLearn workshop*, 2016. 2, 3

[7] Y. Güçlütürk, U. Güçlü, X. Baró, H. J. Escalante, I. Guyon, S. Escalera, M. van Gerven, and R. van Lier. Multimodal first impression analysis with deep residual networks. *IEEE Transactions on Affective Computing*, 2017. 2, 3

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *ArXiv 1512.03385*, 2015. 2

[9] A. C. Little and D. I. Perrett. Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, 98(1):111–126, 2007. 4, 6

[10] V. P. Lopez, B. Chen, A. Clapes, M. Oliu, C. Corneanu, X. Baro, H. J. Escalante, I. Guyon, and S. Escalera. Chalearn lap 2016: First round challenge on first impressions - dataset

and results. In *ChaLearn Looking at People Workshop on Apparent Personality Analysis. ECCV Workshop proceedings, LNCS, Springer, 2016, in press.*, 2016. 2

[11] L. P. Naumann, S. Vazire, P. J. Rentfrow, and S. D. Gosling. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, 35(12):1661–1671, 2009. 6

[12] V. Ponce-Lopez, B. Chen, A. Places, M. Oliu, C. Corneanu, X. Baro, H. Escalante, I. Guyon, and S. Escalera. Chalearn lap 2016: First round challenge on first impressions-dataset and results. In *ChaLearn Looking at People Workshop on Apparent Personality Analysis, ECCVW*, 2016. 2

[13] C. A. M. Sutherland, A. W. Young, and G. Rhodes. Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology*, 2016. 1

[14] S. N. Talamas, K. I. Mavor, and D. I. Perrett. Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PLOS ONE*, 11(2):e0148284, feb 2016. 1

[15] L. Teijeiro-Mosquera, J. I. Biel, J. L. Alba-Castro, and D. Gatica-Perez. What your face vlogs about: Expressions of emotion and big-five traits impressions in youtube. *IEEE Transactions on Affective Computing*, 6(2):193–205, 2015. 2

[16] A. Todorov and J. Porter. Misleading first impressions different for different facial images of the same person. *Psychological Science*, 25(7):1404–17, 2014. 1

[17] A. T. Todorov, C. C. Said, and S. C. Verosky. *Personality Impressions from Facial Appearance*. Oxford Handbook of Face Perception, 2012. 1

[18] C. Ventura, D. Masip, and A. Lapedriza. Interpreting cnn models for apparent personality trait regression. In *CVPRW*, pages 55–63, 2017. 2

[19] M. Walker and T. Vetter. Changing the personality of a face: Perceived big two and big five personality factors modeled in real photographs. *Journal of Personality and Social Psychology*, 110(4):609–624, 2016. 4

[20] Y. Yan, J. Nie, L. Huang, Z. Li, Q. Cao, and Z. Wei. *Exploring Relationship Between Face and Trustworthy Impression Using Mid-level Facial Features*, pages 540–549. Springer International Publishing, 2016. 2

[21] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013. 4

[22] C.-L. Zhang, H. Zhang, X.-S. Wei, and J. Wu. Deep bimodal regression for apparent personality analysis. In *ChaLearn Looking at People Workshop on Apparent Personality Analysis, ECCV Workshop proceedings*, 2016. 2