Article

Personality Trait Analysis in Social Networks Based on Weakly Supervised Learning of Shared Images

Pau Rodríguez [1], Diego Velazquez [2], Guillem Cucurull [1], Josep M. Gonfaus [3], F. Xavier Roca [2], Seiichi Ozawa [4] and Jordi Gonzàlez [2,*]

1   Element AI, Montréal, QC H2S 3G9, Canada; pau.rodriguez@elementai.com (P.R.); gcucurull@cvc.uab.es (G.C.)
2   Computer Vision Center, Campus University Autònoma de Barcelona (UAB), 08193 Bellaterra, Catalonia, Spain; diegoalejandro.velazquez@e-campus.uab.cat (D.V.); xavier.roca@uab.cat (F.X.R.)
3   Visual Tagging Services, Parc de Recerca, Campus UAB, 08193 Bellaterra, Catalonia, Spain; pep.gonfaus@visual-tagging.com
4   Center for Mathematical and Data Sciences, Kobe University, Kobe 657-8501, Japan; ozawasei@kobe-u.ac.jp
*   Correspondence: Jordi.Gonzalez@uab.cat

check for updates

Abstract: Social networks have attracted the attention of psychologists, as the behavior of users can be used to assess personality traits, and to detect sentiments and critical mental situations such as depression or suicidal tendencies. Recently, the increasing amount of image uploads to social networks has shifted the focus from text to image-based personality assessment. However, obtaining the ground-truth requires giving personality questionnaires to the users, making the process very costly and slow, and hindering research on large populations. In this paper, we demonstrate that it is possible to predict which images are most associated with each personality trait of the OCEAN personality model, without requiring ground-truth personality labels. Namely, we present a weakly supervised framework which shows that the personality scores obtained using specific images textually associated with particular personality traits are highly correlated with scores obtained using standard text-based personality questionnaires. We trained an OCEAN trait model based on Convolutional Neural Networks (CNNs), learned from 120K pictures posted with specific textual hashtags, to infer whether the personality scores from the images uploaded by users are consistent with those scores obtained from text. In order to validate our claims, we performed a personality test on a heterogeneous group of 280 human subjects, showing that our model successfully predicts which kind of image will match a person with a given level of a trait. Looking at the results, we obtained evidence that personality is not only correlated with text, but with image content too. Interestingly, different visual patterns emerged from those images most liked by persons with a particular personality trait: for instance, pictures most associated with high conscientiousness usually contained healthy food, while low conscientiousness pictures contained injuries, guns, and alcohol. These findings could pave the way to complement text-based personality questionnaires with image-based questions.

Keywords: sentiment analysis, personality trait analysis; weakly-supervised learning; visual classification; OCEAN model; social networks

1. Introduction

Social media, as a major platform for communication and information exchange, constitutes a rich repository of the opinions and sentiments of 3.6 billion users regarding a vast spectrum of topics. In addition, image sharing on social networks has increased exponentially in the past years; officially

in 2020, Instagram reached 500 million daily active users (1 billion monthly active users) uploading more than 100 million photos and videos per day.

However, current machine learning tools for trends discovery and post analysis are still based mainly on texts, likes behaviour and browsing logs [1–4], though the automatic analysis of such a vast amount of images and videos is becoming a strong feature thanks to the recent advances in deep learning, specifically with the use of Convolutional Neural Networks (CNNs) for image analysis [5]. Such machine learning frameworks have been also applied to psychological assessment [6], first by finding correlations between text and personality [7,8], and later by demonstrating that some image features are related to the personality of users in social networks [9]. A recent overview of the most popular approaches to automated personality detection, various computational datasets, its industrial applications, and state-of-the-art machine learning models for personality detection with specific focus on multimodal approaches vcan be found in [10].

The main hypothesis of this work is that the relationship between text and sentiments conveyed based on personality, as observed by previous researchers like [11], translates well into a relationship between images and personality when we consider the images as being conditioned on specific word use, without any personality annotation. Yarkoni in [11] proved that there exist words that correlate with different personality traits based on statistical evidence (see Table 1 in which 22 words per trait were randomly selected from those identified by Yarkoni [11]). For example, a neurotic personality trait correlates positively with negative emotion words such as 'awful' or 'terrible', whereas an extroverted personality trait correlates positively with words reflecting social settings or experiences like 'bar', 'drinking' and 'crowd'.

Considering this relationship between text and personality, and the fact that posted images have a relationship with their accompanying text, we propose a methodology which, taking advantage of such existing text–personality correlations, exploits the relationship between texts and images in social networks to determine those images most correlated with sentiments conveyed according to personality traits. The final aim was to train a weakly-supervised image-based model for personality and sentiment assessment, that can be used even when textual data is not available.

We considered the human personality characterization called the *Big Five* model of personality for this research [12–14]. The Big Five model is a well-researched personality model, which has been shown to be consistent across age, gender, language and culture [15,16]. In essence, this model distinguishes five main different human personality dimensions: Openness to experience (O), Conscientiousness (C), Extraversion (E), Agreeableness (A) and Neuroticism (N); hence, it is often referred to as *OCEAN*.

These five personality traits have been already related to text [11] and images [17] uploaded by users. Therefore, personality might be an important factor in the underlying distribution of the users' public posts on social media, and thus, it is possible to infer some degree of personality knowledge from such data. However, previous literature on personality assessment from image data has relied on filling personality questionnaires [8,9,11,17], thus slowing down research since it takes a long time to collect data in this way.

In this paper, we train a deep neural network, a Residual Net, that extracts personality representations from images posted in social networks, thus building on previous findings of the association between the big five personality traits and the texts used in social networks. This study goes a step beyond the work of Yarkoni [11], and Segalin et al. [17], showing that it is possible to recover the personality signal from the image domain without explicit image labeling; that is, without conducting intensive personality assessment tests with the authors of the uploaded pictures.

So the innovation presented in this study is to apply the OCEAN model in image training for revealing the personalities of people. Towards this end, a weakly supervised framework for identifying personality traits from images is presented, together with a personality inference model that successfully works on the image domain, which is based on two steps. Firstly, we selected a large set of images whose associated texts were related to the words found to be highly correlated with the big five personality traits. Finally, model prediction was validated with a heterogeneous group

of human subjects by comparing their personality scores with their behaviors in a forced-choice task, where subjects chose between two images that were predicted to have very high and very low scores on a certain personality trait.

In essence, our contributions are:

- A new weakly supervised framework for estimating users personality from images, using the words identified by Yarkoni [11], showing that there exists a correlation between people's personality and their preference for images representing high/low levels of a given trait. The overall idea of predicting personality traits from images without manually labeling them during training is important for a scalable solution.
- A sentiment inference model that successfully works on the image domain, as demonstrated by tests conducted on 280 human subjects (Personality tests were only used to verify our hypotheses, not for training the CNN model.). Quantitative and qualitative results show the evidence that personality traits can be predicted using images.

2. Related Work

The increasing growth and significance of social media in our lives has attracted the attention of researchers, who can use this data in order to infer information about the personality, interests, and behavior of users. Regarding personality, its inference has mainly been based for (i) text uploaded by users, and (ii) uploaded images.

2.1. Text-Based Personality Inference

The relationship between language and personality has been studied extensively. As noted before, Yarkoni [11] performed a large-scale analysis of personality and word use using a large sample of blogs whose authors answered questionnaires to assess their personality. By analyzing the text written by users whose personality was known, the author could investigate the relationship between word use and personality. The results of this analysis concluded that the usage of some specific words was correlated with the personality of the blogs' authors.

Iacobelli et al. [8] used a large corpus of blogs to perform personality classification based on the text of the blogs. They proved that both the structure of the text and the words used were relevant features to estimate personality from the text. Additionally, Oberlander et al. [18] studied whether the personality of blog-authors could be inferred from their personal posts.

In a similar way, Golbeck et al. [7] showed that the personality of users from Twitter could be estimated from their *tweets*, when also taking into account other information such as the number of followers, mentions, or words per *tweet*.

2.2. Image-Based Personality Inference

An early attempt to model personality from images was presented by Steele et al. [19], where they showed that users better agreed with the targets' self-reported personalities when the profile picture was an outdoor picture of a smiling human face.

Cristani et al. [9] proved that there were visual patterns that correlated with the personality traits of 300 Flickr users and, thus, that the personality traits of those users could be inferred from the images they tagged as favorites. To do so, they used aesthetic (colors, edges, entropy, etc) and content features (objects, faces). Guntuku et al. [20] improved the low level features used in a previous work by changing the usual Features-to-Personality (F2P) approach to a two step approach: Features-to-Answers (F2A) + Answers-to-Personality (A2P). Instead of building a model that directly maps features extracted from an image to a personality, with this approach the features are first mapped to the answers of the questionnaire BFI-10 for personality assessment [21]. Then, the answers are mapped to a personality. Besides this two-step approach, they also added new semantic features to extract from the images, like *Black & White vs. Color image*, *Gender identification* and *Scene recognition*.

Segalin et al. [22] proposed a new set of features that better encode the information of the image used to infer the personality of the user who favourited it. They proposed to describe each image using 82 different features, divided into four major categories: *Color*, *Composition*, *Textural Properties* and *Faces*.Their method proved to be suitable to map an image to a personality trait, but it worked better for attributed personality traits rather than self-assessed personality. Following the trend of finding personality-associated features, Ferwerda et al. [23] showed that it is possible to predict the personality of Instagram users by analyzing the color features of their uploaded images. In the same vein, Segalin et al. [24] showed that certain image features, such as color or semantics, or indoors/outdoors, play an important role in determining the personality of Facebook users based on their profile images.

Since a Convolutional Neural Network won the Imagenet competition in 2012 [5], the computer vision field has moved from designing the hand-made image features to learning them in an end-to-end deep learning model. Likewise, the feasibility of using deep learning to automatically learn features that are good to estimate personality traits from pictures has been already demonstrated by the same work by Segalin et al. [17]. In their work, they presented the dataset *PsychoFlickr*, which consists of a collection of images favourited by 300 users from the site *Flickr.com*, with each user tagging 200 images as favorites, adding up to a total of 60,000 images. Additionally, the Big Five traits personality profiles of each user were provided. There were two different versions of the personality profile for each user; one collected through a self-assessment questionnaire answered by the user, and one attributed by a group of 12 assessors who had evaluated the image set of the user. Subsequently, the authors fine-tuned a CNN pre-trained on the large dataset of object classifications known as Imagenet [25], to capture the aesthetic attributes of the images in order to be able to estimate the personality traits associated with those images. For each of the Big Five traits they trained a CNN model with a binary classifier. Then, each different CNN estimate of the images was defined as "high" or "low" for the trait the model was trained for.

The study of the personality conveyed by images has not only been used to infer the personality of users, but also to analyze how brands express and shape their identity through social networks. Ginsberg [26] analyzed the pictures posted to Instagram to interpret the identity of each brand along five dimensions of personality: sincerity, excitement, competence, sophistication, and ruggedness.

Indeed, all previous visual-based approaches take advantage of the many ways in which users interact with images in social networks, such as posting an image, liking it, or commenting on it. Specifically, most of the works described above involved assessing the personality of users based on the images they have liked. For example, the main difference between the work of Segalin et al. [17] and our approach is that, in Segalin's paper, the personality was inferred based on which images were tagged as *favorite*, thus causing it to be a study on the relationship between aesthetic preferences and personality. In contrast, in our case, we explored the image content directly, in particular the content associated with hashtags related to personality traits. The whole procedure is detailed in the next section (the source code used in this paper can be found in this repository: https://github.com/ML-CVC-UKOBE).

3. Methodology

As shown in Yarkoni's work [11], there exists a relationship between the personality of people and the language they use. In other words, the language that we use can reveal our personality traits, so there is statistical evidence that the use of specific words correlate with the personality of online users. Based on this, we designed a set of images, S, conditioned by those words most related to specific personality traits. This can be seen as sampling images from a distribution of images I conditioned on text t, or $S \sim p(I|t)$.

From this set of images S we were able to train a deep learning model that learned to extract a personality representation from a picture, and could use it to automatically infer the personality that the picture conveys. The proposed personality model was built considering a large quantity of

images, tagged with specific personality-related words. These words are those most correlated with each personality trait, as presented by Yarkoni [11]. In Table 1, we showed the words mostly related to each personality trait, which have also been used to identify that set of images for each trait used for training the neural network. Therefore, each image related to a tag will correspond to one of the five personality traits, and within the trait it will represent the *high* or the *low* presence of the trait.

Table 1. List of words highly related to each personality trait, randomly selected from those reported in Yarkoni [11].

Trait	Correlation	Related Words
Openness	High	culture, films, folk, humans, literature, moon, narrative, novel, poet, poetry, sky
	Low	anniversary, detest, diaper, hate, hatred, hubby, implore, loves, prayers, thankful, thanks
Conscientiousness	High	achieved, adventure, challenging, determined, discipline, persistence, recovery, routine, snack, vegetables, visit
	Low	bang, bloody, boring, deny, drunk, fool, protest, soldier, stupid, swear, vain
Extraversion	High	bar, concert, crowd, dancing, drinking, friends, girls, grandfather, party, pool, restaurant
	Low	blankets, books, cats, computer, enough, interest, knitting, lazy, minor, pages, winter
Agreeableness	High	afternoon, beautiful, feelings, gifts, hug, joy, spring, summer, together, walked, wonderful
	Low	assh*le, bin, cost, drugs, excuse, harm, idiot, p*rn, sexual, stupid, violence
Neuroticism	High	annoying, ashamed, awful, horrible, lazy, sick, stress, stressful, terrible, upset, worse
	Low	completed, county, ground, later, mountain, oldest, poem, road, southern, sunset, thirty

As the aim of this paper was to evaluate whether there is any of the author's personality information embedded in real world images, we have considered an image-centered social network, Instagram, which is widely used for user behavior analysis [26–28]. For example, Hu et al. [27] showed that the pictures posted to Instagram can be classified into eight main categories, and that users can be divided into five different groups, depending on what kind of pictures they post. These eight main picture categories are: friends, food, gadgets, captioned pictures, pets, activity, selfies, and fashion. The differences on the type of images posted can be influenced by the city of the users, [29] or their age [30].

In order to determine the set of images, we used the words from Table 1 to query images. For each personality trait, 22 words were used (11 for each component), and about 1100 images were selected for each word. This selection was done by matching the hashtag associated with the image with the words listed in each column, which contained 10 images selected by matching their hashtags with the list of words in Table 1. As a result, the number of training images was balanced within the five personality traits, and also with the used tags, thus each trait was trained with around 24,000 images.

In Figure 1 we show some ten random samples of each personality trait obtained with the procedure described above. As shown, despite the huge intra-class and inter-class variabilities of the images associated with each of the personality traits, there are consistence between the images of the same class, whose hashtags are related to the words suggested by Yarkoni [11].
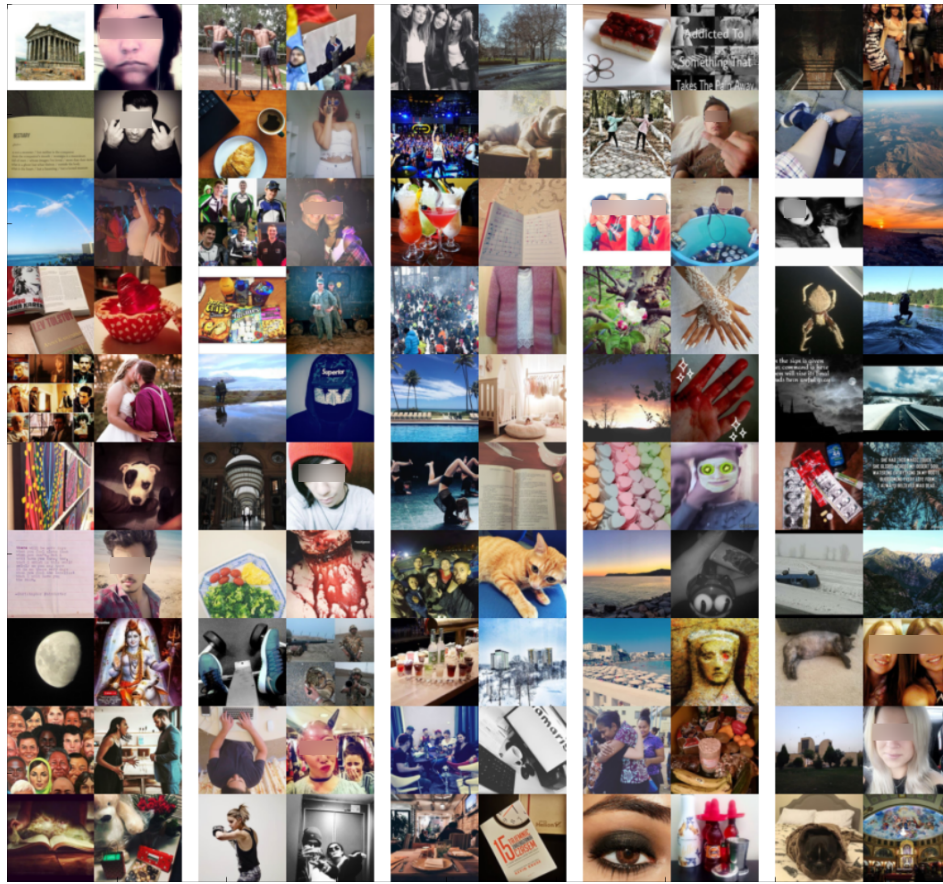
Figure 1. Image samples. Each column contains 10 images selected by matching their hashtags with the list of words listed in Table 1. From left to right: High openness, Low openness; High conscientiousness, Low conscientiousness; High extraversion, Low extraversion; High agreeableness, Low agreeableness ; High neuroticism, Low neuroticism.

Building the Personality Model

Once the procedure had been defined to determine which set of images S was most related to each personality trait, we modelled this relationship between images and personality. In other words, we next built a model which was able to predict the personality trait corresponding to an image, given the training set described in the previous section.

In this work, we have used a CNN model that maps an input image to a desired output by learning a set of parameters that produce a good representation of the input. This model is hierarchical; it consists of several layers of feature detectors that build a hierarchical representation of the input, from local edge features to abstract concepts. The final layer consists of a linear classifier, which projects the last layer features into the label space. Let x be an input image and $f(x;\theta)$ a parametric function that maps this input to an output, where θ are the parameters. The neural network model is a hierarchical combination of computation layers:

$$f(x;\theta) = f^N(f^{N-1}(...f^2(f^1(x;\theta_1);\theta_2);\theta_{N-1});\theta_N) \tag{1}$$

where N is the number of layers in the model, and each computation layer corresponds to a non-linear function f^n [31] with its own parameters θ_n. We find θ by empirical risk minimization:

$$\theta = \arg\min_\theta \mathcal{L}(y,\hat{y}), \tag{2}$$

where $\mathcal{L}$ is the cross-entropy loss function, y is the output of the model defined as $y = f(x; \theta)$, and $\hat{y}$ is the correct output for input x. This function is minimized iteratively by means of Stochastic Gradient Descent (SGD), and the process of minimizing the loss function over a set of images S is referred as *training* the model. In this work, we used a CNN model [5,32], since CNNs are especially suited for 2D data. After training, the output of the CNN for an image is a vector of scores for each of the personality traits.

For predicting the high and low scores for each of the personality traits, we propose an all-in-one model: we used the same CNN for all five traits, but using five different classifiers on the last layer, one for each of the Big Five traits. Each output layer was independent of the others, and consisted of a binary classifier like that described before, which has its own loss function.

There are some common CNN architectures [5,33–35] that are well established for computer vision tasks. In our experiments, we have used a Residual Network presented by He et al. [35] (*ResNet*), which uses residual connections between the network layers, allowing for easier optimization, and thus allowing us to increase the network depth. In the experiments, we used a ResNet of 50 layers, while replacing the last layer to contain the required amount of outputs for the two configurations explained previously.

We tested two different options for initializing the network's weights: (i) randomly sampling from a Gaussian distribution, and (ii) fine-tuning, which consisted of initializing the network with the weights learned for another task. In this case, we initialized the network with the weights of a model trained on ImageNet [36]. The fine-tuning approach has been proved by Oquab et al. [37] to be useful to train neural networks in small datasets, with superior performance. The idea behind this approach is that the network first learns how to extract good visual features in a large dataset, and then uses this features to learn a classifier in a smaller dataset.

We trained the models with Caffe [38] on a NVIDIA GTX 770 with 4GB of memory. The models were optimized with SGD with a momentum of 0.9 and batch size of 128 for Alexnet and ResNet, respectively. The learning rate was set to 0.01 when training a network from scratch, and to 0.001 when fine-tuning, increasing the learning rate by a factor of 10 at the new layers. Additionally, during the training stage, we randomly applied horizontal mirroring to the images and cropped a random patch of 224×224 pixels of the original 256×256 images. The only pre-processing of the images was the subtraction of the training set mean. A random split of the dataset was used to divide the images in non-overlapping training and testing sets; 80% of the images were used for training and 20% for testing.

4. Experimental Evaluation of Psychological Traits

The performances on personality recognition for each trait are shown in Figure 2. As it can be seen, the Extraversion trait obtained the best scores, followed by Conscientiousness. The dashed line in the ROC curve represents a hypothetical random classifier with an AUC of 0.5. As it can be seen, the ROC curves for each personality trait are far from being random, indicating that a relationship between text and images exists, and that the model is able to learn a mapping from image to personality. These results show the personality classifiers share feature extraction layers, and each classifier contributes to the learning.

In order to demonstrate the feasibility of the proposed weakly supervised learning scheme, we tested our model predictions on a random pool of 280 adult human subjects recruited online, of different ages (all less than 40 years), gender (90% male), and nationality. Each subject completed first a standard Big 5 personality assessment questionnaire [39].

(**a**) ROC curves for all traits   (**b**) PR curves for all traits
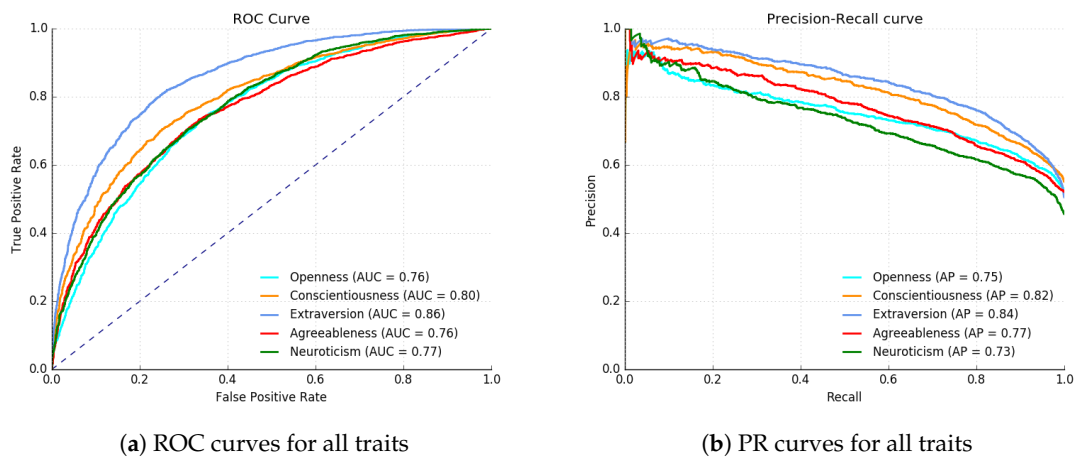
Figure 2. Receiver operating characteristic and Precision-recall curves for all the personality traits. The ROC plot also shows the Area under the curve for each trait, and the PR plot shows the Average precision for each trait.

4.1. Comparing Image Classification with Human-Based Classification

The questionnaire consisted of fifty text-based questions that the user answered with a score from one (strongly disagree) to five (strongly agree), and the outcome was a vector of five numbers from -1 to 1 (after normalization), corresponding to each of the five personality traits in the *OCEAN* model. The distribution of the answers is shown in Figure 3.

Subsequently, subjects were presented with two opposite images for each trait, as predicted by our model. That is, pairs of images were selected from a random set of images also associated with the hashtags listed in Table 1, but not used for training. The use of the CNN trained as described in the Methodology section was aimed at ranking those test images, according to their high or low levels of each of the five traits.

Therefore, at the image-based questionnaire step, the user had to choose between an image predicted with a high conscientiousness score and one with very low conscientiousness, and this was repeated for each trait. Table 2 shows samples of the images shown to users. For each experiment, the ordering of the images was sampled from a random uniform distribution to mitigate possible effects of the interface on the user choices.

In Figure 4, we show the Pearson correlations between the score of each of the questionnaire-assessed traits, and the images chosen by the users. Given the test output for a personality trait, we were able to predict the image choice associated with that trait better than for any of the other traits. This indicates that the CNN was in most cases able to predict which images were the most associated with a particular personality trait. Interestingly, it can be seen some confusion can be seen between Agreeableness and Extraversion. Likewise, in the works of Goldberg [14] and Segalin et al. [24], it was shown that extraversion and agreeableness share similar correlations for different image features, see Figure 5. The possible implications of these findings are outside of the scope of this work, and we propose they be considered in future research.

Detailed results are shown in Figure 6. The exact image choice was compared with the personality test score, showing that given the test score for a given trait (positive or negative), users chose the image that our model predicted as being most associated with the trait score. In Figure 6a, we show the precision score of a user choosing the image we predicted as most associated with a certain trait:

$$Precision(trait) = \frac{|test = trait \cap user = trait|}{|user = trait \cup user \neq trait|},$$

where *user* is the user choice (for instance, the image associated with *O+*), and *test* is the test score for the given *trait*. Figure 6b shows the spearman correlation coefficient between the user image

choice and the questionnaire score. As shown, for each trait score, users tend to choose the image we predicted to be most associated with the that trait score, thus providing more evidence supporting the proposed framework.

Table 2. Sample of the images and questions used to assess users' personality. None of these images were used for CNN training.
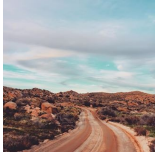
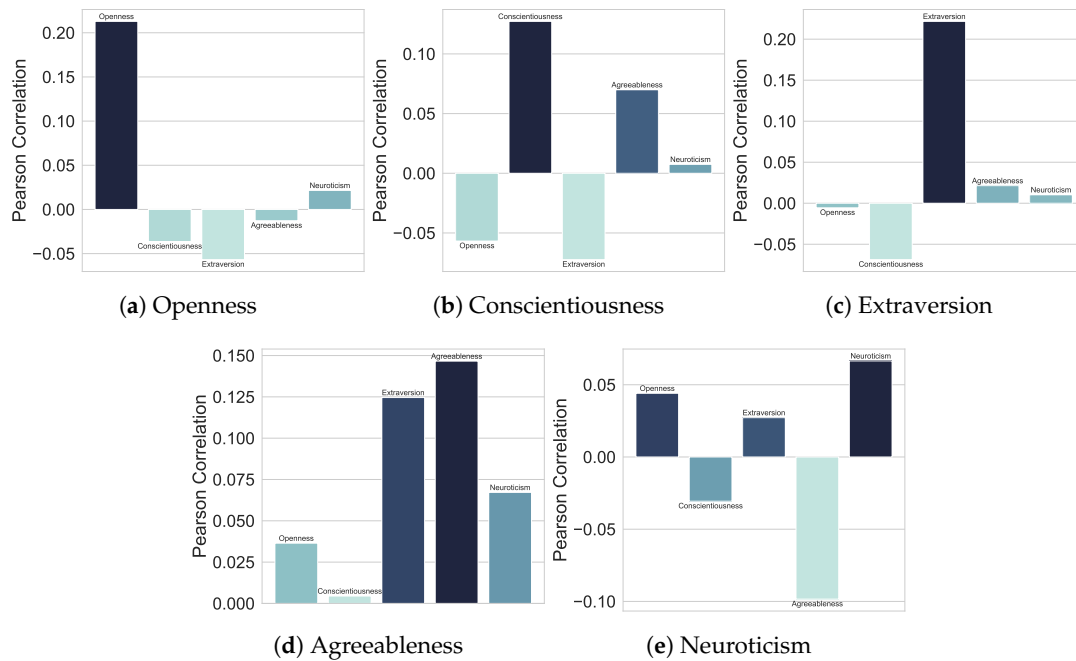| Trait		Textual Questions	Visual Question Samples
O	High	• I have a rich vocabulary • I have a vivid imagination • I have excellent ideas	
	Low	• I have difficulty understanding abstract ideas • I am not interested in abstract ideas. • I do not have a good imagination.	
C	High	• I am always prepared • I pay attention to details • I like order	
	Low	• I leave my belongings around • I make a mess of things • I often forget to put things back in their proper place	
E	High	• I am the life of the party • I feel comfortable around people • I start conversations	
	Low	• I don't talk a lot. • I keep in the background • I have little to say.	
A	High	• I am interested in people • I sympathize with others' feelings • I take time out for others.	
	Low	• I feel little concern for others • I insult people. • I am not interested in other people's problems.	

Table 2. *Cont.*

| Trait		Textual Questions	Visual Question Samples
N	High	• I get stressed out easily. • I worry about things. • I am easily disturbed.	
	Low	• I am relaxed most of the time. • I seldom feel sad. • I remain calm in tense situations	
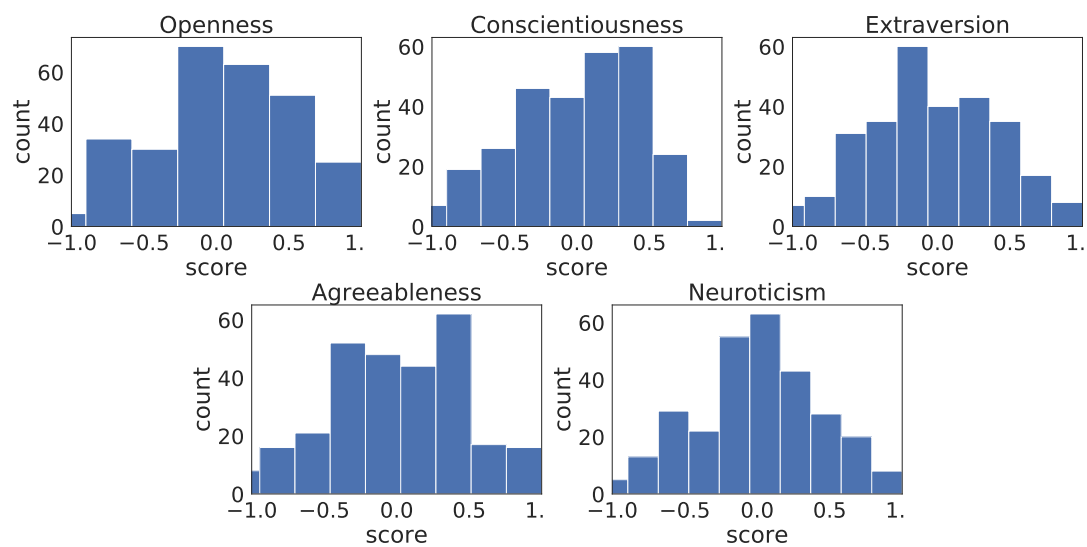
Figure 3. Histograms with the distribution of scores on the personality questionnaires for each of the big five personality traits. Scores have been normalized to be between −1 and 1.
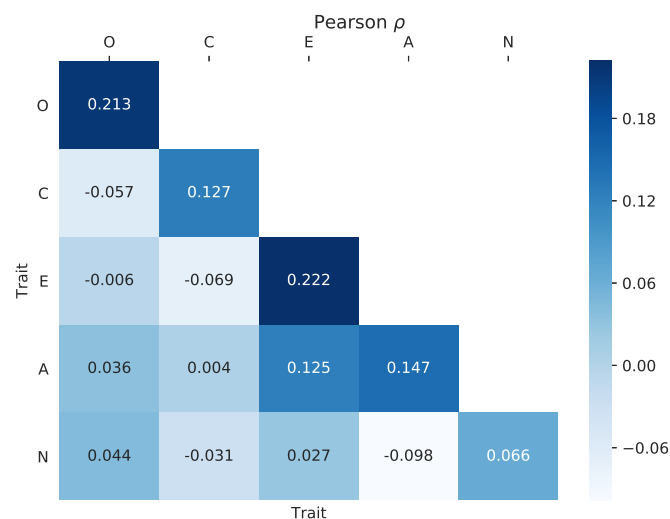
Figure 4. Pearson correlation coefficient (ρ) between the image choice and the personality score in the questionnaires. Knowing the test score for a personality trait allows prediction of the image choice for the chosen personality trait. Thus, our model is able to successfully predict images that a person with a given level of a trait will like.

(**a**) Openness   (**b**) Conscientiousness   (**c**) Extraversion

(**d**) Agreeableness   (**e**) Neuroticism

Figure 5. Pearson correlation coefficient between user questionnaire results (subfigures) and our model predictions on the user-chosen images (columns). Images predicted to be associated with certain personality trait are the most highly correlated with the corresponding questionnaire results.

(**a**) Precision matrix   (**b**) Spearman correlation matrix
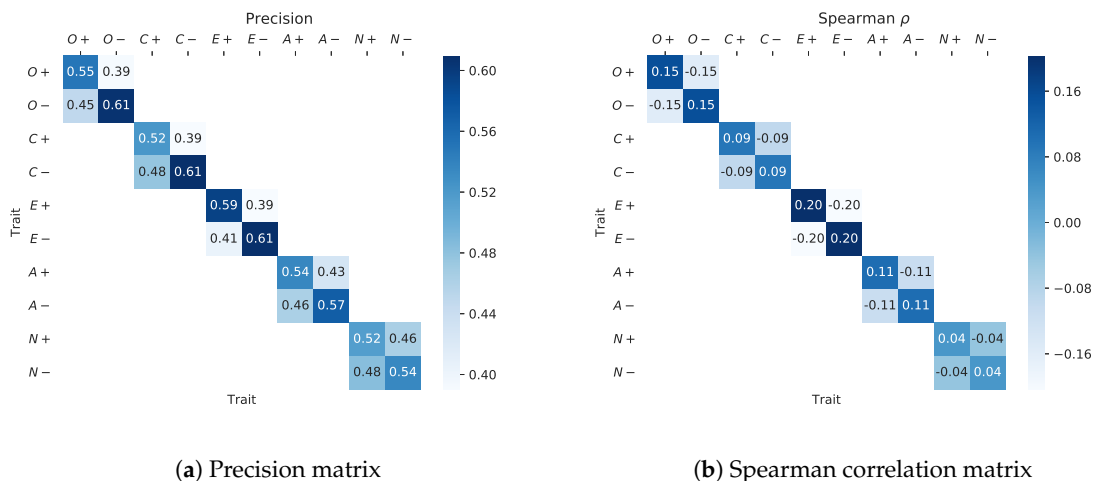
Figure 6. Precision (**a**) and Spearman Correlation (**b**) matrices between the test-assessed personality traits and the image choices. Users with a given level of a trait tended to like the same model-selected images.

4.2. Qualitative Results

To get a better insight into what the new deep features are detecting, we visualized and analyzed the images that maximally activated the output of the network for each personality trait. Namely, in order to know which images better represented each personality trait, we found those pictures that maximally activated a specific output of our model. Similarly to Girshick et al. [40], we input all the images into our model, inspected the activation values of a specific neuron, and looked for the images that produce these maximum activations.

In our case, we inspected the output units associated with each personality trait. For example, by looking at the output of our Extraversion classifier for each of the images, we could know which ones were classified as High Extraversion with more confidence.

In Figure 7, the most representative images of the High and Low scores for each trait are shown. Among the pictures that maximally activated the High Openness trait we can see pictures of books,

the moon, and the sky, while for Low Openness the most relevant pictures were love related. For High Conscientiousness most of the images were photographs of food, especially healthy food, whereas for Low Conscientiousness we mostly see pictures of alcohol, guns, and people. In Extraversion, there was a clear distinction between the images that maximally activated the High and Low outputs. The High Extraversion output was mostly activated by pictures of a lot of people, whereas the Low Extraversion output reacted to cats, books, and knitting images. In High Agreeableness, we mostly see flower pictures, whereas the Low score responds to pictures with text and naked torsos. Lastly, in the Neuroticism trait we observe that the High score was maximally activated by pets, whereas for the Low score we see pictures of landscapes and sunsets.

(**a**) High Openness　　　　(**b**) Low Openness　　　(**c**) High Conscientiousness (**d**) Low Conscientiousness

(**e**) High Extraversion　　　(**f**) Low Extraversion　　　(**g**) High Agreeableness　　(**h**) Low Agreeableness

(**i**) High Neuroticism　　　(**j**) Low Neuroticism

Figure 7. Images that maximally activate the OCEAN traits.

5. Conclusions

We proposed a weakly supervised learning framework for obtaining image-based personality data without the use of personality questionnaires, thus providing new tools to deal with the ground-truth annotation bottleneck. The underlying hypothesis of this work is that when a user posts a picture to a social network, the picture is not expressing everything the author has in mind, but only a specific message or mood of the author. Therefore, a picture does not describe the whole personality of the user, but a portion of it. Therefore, in order to obtain an estimation of the whole personality profile of the users, one could analyze all the different images posted by them, because each image conveys only partial information regarding their personalities.

In particular, we showed that, given images uploaded by users who used words most used by persons with a particular personality trait [11], it is possible to predict which images persons with a specific trait are more likely to select, as we confirmed by conducting personality tests with 280 subjects. Thus, images and text are correlated, which can be explained if both depend on a third variable, which is personality. In this study, we did not recover the full spectrum of personality traits of one user, but we inferred the most prominent personality trait that a single image conveys.

There are inherent limitations in this study, which pave the way of the future work. First, the use of pairwise choice data to assess the model performance could be problematic. Since the output space of the CNN is 10-D (the positive and negative categories of the five personality traits) and not 2-D (the positive and native categories of one personality trait), in the future we will ask each participant to choose one image from ten alternatives (each of the ten alternatives will be randomly selected from each of the ten personality trait categories respectively). This way, for each participant, one will see if his/her most prominent personality trait (or trait category) indeed corresponds to he/she choosing the image that our model predicts to score the highest on the same trait (or trait category).

Secondly, it is not evident whether using Residual Nets is optimal for this task. Recently, Zhu et al. in [41] proposed an attention-based network to infer personality traits. In their work, the authors use Class Activation Maps (CAMs) to find relevant regions for personality traits in the image. We believe that including similar attention based approach can improve the performance of our work..

Finally, we will analyze how this proposed model generalizes on images uploaded in social networks without the hashtags used in this work. In this way, we plan to evaluate whether that the resulting classifier is able to not only determine high or low personality traits, but also visual topics and regions strongly correlated to such personality traits.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area under the curve
CNN	Convolutional Neural Network
OCEAN	Openness to experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism
PR	Precision-recall
ROC	Receiver operating characteristic

References

1. Camacho, D.; Panizo-LLedot, A.; Bello-Orgaz, G.; Gonzalez-Pardo, A.; Cambria, E. The Four Dimensions of Social Network Analysis: An Overview of Research Methods, Applications, and Software Tools. *Inf. Fusion* **2020**, *63*, 88–120. [CrossRef]
2. Goyal, P.; Kaushik, P.; Gupta, P.; Vashisth, D.; Agarwal, S.; Goyal, N. Multilevel Event Detection, Storyline Generation, and Summarization for Tweet Streams. *IEEE Trans. Comput. Soc. Syst.* **2019**, *7*, 8–23. [CrossRef]

3. Li, L.; Zhang, Q.; Wang, X.; Zhang, J.; Wang, T.; Gao, T.L.; Duan, W.; Tsoi, K.K.f.; Wang, F.Y. Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on weibo. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 556–562. [CrossRef]

4. Madisetty, S.; Desarkar, M.S. A neural network-based ensemble approach for spam detection in Twitter. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 973–984. [CrossRef]

5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.

6. Cambria, E. Affective computing and sentiment analysis. *IEEE Intell. Syst.* **2016**, *31*, 102–107. [CrossRef]

7. Golbeck, J.; Robles, C.; Edmondson, M.; Turner, K. Predicting personality from twitter. In Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, USA, 9–11 October 2011; pp. 149–156.

8. Iacobelli, F.; Gill, A.J.; Nowson, S.; Oberlander, J. Large scale personality classification of bloggers. In *Affective Computing and Intelligent Interaction*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 568–577.

9. Cristani, M.; Vinciarelli, A.; Segalin, C.; Perina, A. Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In Proceedings of the 21st ACM International Conference on Multimedia, ACM, Barcelona, Spain, 21–25 October 2013; pp. 213–222.

10. Mehta, Y.; Majumder, N.; Gelbukh, A.; Cambria, E. Recent trends in deep learning based personality detection. *Artif. Intell. Rev.* **2020**, *53*, 1313–1339. [CrossRef]

11. Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *J. Res. Personal.* **2010**, *44*, 363–373. [CrossRef]

12. Digman, J.M. Personality structure: Emergence of the five-factor model. *Annu. Rev. Psychol.* **1990**, *41*, 417–440. [CrossRef]

13. Barrick, M.R.; Mount, M.K. The big five personality dimensions and job performance: A meta-analysis. *Pers. Psychol.* **1991**, *44*, 1–26. [CrossRef]

14. Goldberg, L.R. An alternative "description of personality": The big-five factor structure. *J. Personal. Soc. Psychol.* **1990**, *59*, 1216. [CrossRef]

15. McCrae, R.R.; John, O.P. An introduction to the five-factor model and its applications. *J. Personal.* **1992**, *60*, 175–215. [CrossRef] [PubMed]

16. Schmitt, D.P.; Allik, J.; McCrae, R.R.; Benet-Martínez, V. The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *J. Cross-Cult. Psychol.* **2007**, *38*, 173–212. [CrossRef]

17. Segalin, C.; Cheng, D.S.; Cristani, M. Social profiling through image understanding: Personality inference using convolutional neural networks. *Comput. Vis. Image Underst.* **2016**, *156*, 34–50. [CrossRef]

18. Oberlander, J.; Nowson, S. Whose thumb is it anyway? Classifying author personality from weblog text. In Proceedings of the COLING/ACL on Main Conference Poster Sessions, Association for Computational Linguistics, Sydney, Australia, 17–18 July 2006; pp. 627–634.

19. Steele, F.; Evans, D.C.; Green, R.K. Is Your Profile Picture Worth 1000 Words? Photo Characteristics Associated with Personality Impression Agreement. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009.

20. Guntuku, S.C.; Roy, S.; Lin, W. Personality Modeling Based Image Recommendation. In *Proceedings of the International Conference on Multimedia Modeling*; Springer: Cham, Switzerland, 2015; pp. 171–182.

21. Rammstedt, B.; John, O.P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. Res. Personal.* **2007**, *41*, 203–212. [CrossRef]

22. Segalin, C.; Perina, A.; Cristani, M.; Vinciarelli, A. The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits. *IEEE Trans. Affect. Comput.* **2016**, *8*, 268–285. [CrossRef]

23. Ferwerda, B.; Schedl, M.; Tkalcic, M. Using instagram picture features to predict users' personality. In *Proceedings of the International Conference on Multimedia Modeling*; Springer: Cham, Switzerland, 2016; pp. 850–861.

24. Segalin, C.; Celli, F.; Polonio, L.; Kosinski, M.; Stillwell, D.; Sebe, N.; Cristani, M.; Lepri, B. What your Facebook profile picture reveals about your personality. In Proceedings of the 2017 ACM on Multimedia Conference, Silicon Valley, CA, USA, 23–27 October 2017; pp. 460–468.

25. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR09, Miami, FL, USA, 20–25 June 2009.

26. Ginsberg, K. Instabranding: Shaping the personalities of the top food brands on instagram. *J. Undergrad. Res.* **2015**, *6*, 78-91.

27. Hu, Y.; Manikonda, L.; Kambhampati, S. What We Instagram: A First Analysis of Instagram Photo Content and User Types. In Proceedings of the ICWSM, Ann Arbor, MI, USA, 1–4 June 2014.

28. Souza, F.; de Las Casas, D.; Flores, V.; Youn, S.; Cha, M.; Quercia, D.; Almeida, V. Dawn of the selfie era: The whos, wheres, and hows of selfies on Instagram. In Proceedings of the 2015 ACM on Conference on Online Social Networks, Palo Alto, CA, USA, 2–3 November 2015; pp. 221–231.

29. Hochman, N.; Schwartz, R. Visualizing instagram: Tracing cultural visual rhythms. In Proceedings of the Workshop on Social Media Visualization (SocMedVis) in Conjunction with the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM–12), Dublin, Ireland, 4–7 June 2012; pp. 6–9.

30. Jang, J.Y.; Han, K.; Shih, P.C.; Lee, D. Generation like: Comparative characteristics in Instagram. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 4039–4042.

31. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 17 November 2020).

32. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

34. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

36. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

37. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1717–1724.

38. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv* **2014**, arXiv:1408.5093.

39. Goldberg, L.R. Possible Questionnaire Format for Administering the 50-Item Set of IPIP Big-Five Factor Markers. *Psychol. Assess.* **1992**, *4*, 26–42. [CrossRef]

40. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

41. Zhu, H.; Li, L.; Jiang, H.; Tan, A. Inferring Personality Traits from Attentive Regions of User Liked Images Via Weakly Supervised Dual Convolutional Network. *Neural Process. Lett.* **2020**, *51*, 2105–2121. [CrossRef]