Integrating Vision and Language in Social Networks for Identifying Visual Patterns of Personality Traits

Pau Rodriguez, Jordi Gonzàlez, Josep M. Gonfaus, and F. Xavier Roca

***Abstract*—Social media, as a major platform for communication and information exchange, is a rich repository of the opinions and sentiments of 2.3 billion users about a vast spectrum of topics. In this sense, user text interactions are widely used to sense the whys of certain social user's demands and cultural- driven interests. However, the knowledge embedded in the 1.8 billion pictures which are uploaded daily in public profiles has just started to be exploited. Following this trend on visual-based social analysis, we present a novel methodology based on neural networks to build a combined image-and-text based personality trait model, trained with images posted together with words found highly correlated to specific personality traits. So, the key contribution in this work is to explore whether OCEAN personality trait modeling can be addressed based on images, here called *MindPics*, appearing with certain tags with psychological insights. We found that there is a correlation between posted images and the personality estimated from their accompanying texts. Thus, the experimental results are consistent with previous cyber-psychology results based on texts, suggesting that images could also be used for personality estimation: classification results on some personality traits show that specific and characteristic visual patterns emerge, in essence representing abstract concepts. These results open new avenues of research for further refining the proposed personality model under the supervision of psychology experts, and to further substitute current textual personality questionnaires by image-based ones.**

***Index Terms*—Personality trait analysis, deep learning, visual classification, OCEAN model, social networks.**

I. Introduction

Image sharing in social networks has increased exponentially in the past years. Officially, there are 600 million Instagrammers uploading around 100 million photos and videos per day. Although the analysis of trends, topics and brands in social networks is mostly based solely on texts, the analysis of such a vast number of images is starting to play an important role for understanding and predicting human decision making, while becoming essential for digital marketing, and customer understanding, among others. Indeed, previous works have proven the relation between text and the personality of the authors [1], [2], and recent studies have also shown that some image features can be related to the personality of users in social networks [3].

The main hypothesis of this work is that the relation between text and personality observed by researchers like Yarkoni [4] translates well into a relation between images and personality when we consider the images conditioned on specific word use, without any personality annotation. In his work, Yarkoni proved that there exist words that correlate with different personality traits with statistical evidence, see Table I. For example, a neurotic personality trait correlates positively with negative emotion words such as 'awful' or 'terrible', whereas an extroverted person correlates positively with words reflecting social settings or experiences like 'bar', or 'crowd'. Considering this proven relation between text and personality, and the fact that posted images have a relation with their accompanying texts, we propose a methodology which, taking advantage of such existing text-personality correlation, exploits the relation between texts and images in social networks to determine those images most correlated with personality traits. The final aim is to use this set of images to train a personality model with similar performances than previous works based on texts or images alone.

In the computer vision community, the relationship between language and images has been exploited to automatically generate textual descriptions from pictures [5]–[7]. Thus, automatic captioning can be understood as a sampling process of words from a text distribution t given an image I, or $p(t|I)$. In this paper we aim at the opposite: we want to determine $p(I|t)$, that means, we look for those images most correlated with those words most associated to particular personality traits. So, we aim at determining the potential relation between personality and images using a state-of-the-art deep neural network. High classification results will suggest that the psychological traits found in the text are still present in the image.

In our work, the human personality characterization called the Big Five model of personality is considered [8]–[10]. The Big Five model is a well-researched personality description, which has been proven to be consistent across age, gender, language and culture [11], [12]. 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 as OCEAN and characterized by the following features:

- *Openness*: Appreciate art and ideas, imaginative, aware of feelings. People with this trait tend to have artistic interests and have a certain level of intellectuality
- *Conscientiousness*: Disciplined, dutiful, persistent, compulsive and perfectionist as opposite to spontaneous and

Manuscript received December 9, 2018; revised February 23, 2018. This work was supported by the Spanish project TIN2015-65464-R (MINECO/FEDER), the 2016FI B 01163 grant of Generalitat de Catalunya, and the COST Action IC1307 iV&L Net.

P. Rodriguez, F. Xavier Roca, and J. Gonzàlez are with the Computer Vision Center, Edifici O, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Catalonia Spain (e-mails: prodriguez, xavir, poal@cvc.uab.es).

J. M. Gonfaus is with Visual Tagging Services S.L., Parc de Recerca UAB, Edifici Eureka, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Catalonia Spain (e-mail: pep.gonfaus@visual-tagging.com).

doi: 10.18178/ijssh.2019.9.1.981

impulsive. People with this trait tend to strive for some- thing, and to be hard-workers and organized.

- *Extraversion*: Warm, assertive, action-oriented, and thrill- seeking. Individuals with high levels of extraversion tend to be friendly, sociable, cheerful and fond of being in company with other people.
- *Agreeableness*: Compassionate, cooperative, considerate. Agreeable people tend to be trusty, modest and optimistic.
- *Neuroticism*: Emotional instability, anxious, hostile, prone to depression. Neurotics tend to be frustrated, anxious and experience negative emotions.

The five personality traits have already been related to text [4] and images [13] uploaded by users. Therefore, personality might be an important factor in the underlying distribution of the user's public posts in social media, and thus, it is possible to infer some degree of personality knowledge from such data. In this work we go a step beyond the works in [4], and [13], showing that personality remains invariant to changes from the text domain to the image domain. Concretely, our contributions are:

- A new framework for identifying visual patterns or images based on personality traits, called MindPics, accessed using a refined set of the tags proposed in [4].
- A personality inference model that uses MindPics, as a proof that personality remains invariant across textual and visual domains.

II. Related Work

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

A. Text-Based Personality Inference

The relationship between language and personality has been studied extensively. As commented before, Yarkoni et al. [4] performed a large-scale analysis of personality and word use in a large sample of blogs whose authors answered questionnaires to assess their personality, see Table I.

TABLE I: List of Words Highly Related to Each Personality Trait, Extracted From [4]

Trait	Related Words
Openness	culture, films, folk, humans, literature, moon, narrative, novel, poet, poetry, sky
Conscientiousness	achieved, adventure, challenging, determined, discipline, persistence, recovery, routine, snack, vegetables, visit
Extraversion	bar, concert, crowd, dancing, drinking, friends, girls, grandfather, party, pool, restaurant
Agreeableness	afternoon, beautiful, feelings, gifts, hug, joy, spring, summer, together, walked, wonderful
Neuroticism	annoying, ashamed, awful, horrible, lazy, sick, stress, stressful, terrible, upset, worse

This way, by analyzing the text written by users whose personality is known, the author could investigate the relation between word use and personality. The results of the analysis concluded that the usage of some specific words is correlated with the personality of the blogs' authors.

Iacobelli *et al*. [2] 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 are relevant features to estimate the personality from text. Also, Oberlander *et al*. [14] studied if the personality of blog-authors could be inferred from their personal posts.

In a similar way, Golbeck et al. [1] showed that the personality of users from Twitter could be estimated from their tweets taking into account also other information as the number of followers, mentions or words per tweet.

B. Image-Based Personality Inference

An early attempt to model personality from images was presented in Cristiani et al. [3], which proved that there are visual patterns that correlate with the personality traits of 300 Flickr users and thus, that personality traits of those users could be inferred from the images they tagged as favorite. To do so, aesthetic (colors, edges, entropy, etc) and content features (objects, faces).

Guntuku *et al*. [15] improved the low-level features used in 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 [16]. Then, the answers are mapped to a personality. Besides this two-step approach, they also add new semantic features to extract from the images, like Black & White vs. Color image, Gender identification and Scene recognition.

Later, Segalin *et al*. [17] proposed a new set of features that better encode the information of the image used to infer the personality of the user who favorited it. They proposed to describe each image with 82 different features, divided in 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.

Since a Convolutional Neural Network (CNN) won the *Imagenet* competition in 2012 [18] the computer vision field has moved from designing the hand-made image features to learn them in an end-to-end deep learning model. Likewise, the feasibility of deep learning to automatically learn features that are good to estimate personality traits from pictures have been already proven by the same work of Segalin et al. [13].

In their work, Segalin et al. presented the dataset *PsychoFlickr*, which consists of a collection of images favorited by 300 users from the site Flickr.com, each user tagging 200 images as favorite, adding up to a total of 60,000 images. Additionally, the Big Five traits personality profile of each user are provided. There are 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 tune a CNN pre-trained on the large dataset of object classification *Imagenet* 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 estimates if the images are related for the trait the model has been 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 et al. [19] analyzed the pictures posted in Instagram to interpret the identity of each brand along five dimensions of personality: sincerity, excitement, competence, sophistication, and ruggedness.

C. Integrating Text and Image for Personality Inference

Table II contains the recognition accuracy of each of the OCEAN traits for 5 different methods based on text or images. In this paper, we want to determine whether the correlations found between personality and texts or images separately, also holds when combining image and word use.

TABLE II: CLASSIFICATION ACCURACIES (%) REPORTED IN THE LITERATURE USING TEXTS OR IMAGES

	Word use only		Image use only	
	Golbeck [1]	Iacobelli [2]	Segalin [13]	Guntuku [16]
Openness	75.50	84.36	61.00	66.10
Conscientiousness	61.70	79.18	67.00	70.50
Extraversion	58.60	71.68	65.00	69.70
Agreeableness	69.70	78.31	64.00	72.30
Neuroticism	42.80	70.51	69.00	61.50
Average	61.66	76.80	65.20	68.02

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 consist on assessing the personality of users based on the images they have liked. For example, the main difference between [13] and our approach is that in such paper, the personality is inferred based on which images have been tagged as favorite, thus becoming a study on the relation between aesthetic preferences and personality. In contrast, in our case, we explore directly the images shared by users based on accompanying texts strongly related to personality traits, so here the relationship between images and personality arises from the mere act of posting a picture in a social network as a process of communication with others. The whole procedure is detailed next.

III. METHODOLOGY

As proven by Yarkoni *et al.* [4] 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 correlates with the personality of online users. Based on that, we design a set of images *S* conditioned by those words most related to specific personality traits, see the whole process in Figure 1. This can be seen as sampling images from a distribution of images *I* conditioned on text *t*.

For each trait of the Big Five personality traits defined before we have selected the list of words that correlate most positively or negatively with the trait, as suggested by [4]. The positively correlated words have been used to identify those images most associated to the strong presence of a trait, and the negatively correlated words are used to determine those images most associated to its absence.

From this set of images we will train a deep learning model that learns to extract a personality representation from a picture, and use it to automatically infer the personality that the picture conveys.

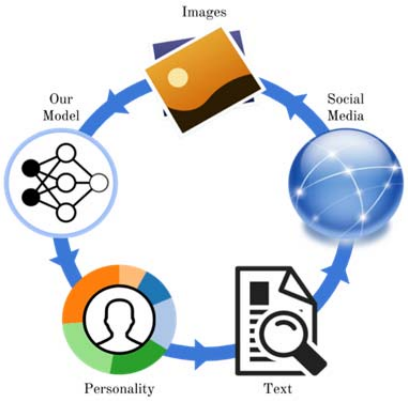

Fig. 1. Modular scheme of the methodology presented in this paper: those textual words most correlated with personality traits are used as hashtags of those images uploaded in Social Networks. These images will be used for training a Neural Network which is able to predict the personality trait based on images only.

A. Finding MindPics in Social Networks

Based on the aforementioned relationship between text and personality, the proposed personality model is built considering a large quantity of images, called *MindPics*, tagged with specific personality-related words. These words are the most correlated ones with each personality trait, as presented in [4]. In Table I, we showed the words mostly related to each personality trait, which have been also used to identify that set of images for each trait used for training the neural network. So, each image related to a tag will correspond to one of the five personality traits, and within the trait it will represent the high presence of such trait.

As the aim of this paper is to evaluate whether there is any of the author's personality information embedded in real world images, we have considered images posted in Instagram [19]– [21]. In this social network the users take and share images by posting them together with words or hashtags. To build the *MindPics* dataset, we first crawl a large collection of publicly- shared photos using the Instagram API.

The reason of choosing Instagram as the source of selecting the images for training the Neural Network is threefold:

(i) images are the main content of Instagram instead of text. Contrary to more text-based social networks such as Twitter, most of the content and information shared by the Instagram user is conveyed in the image, so it is reasonable to think that such images embed personality, up to some extent;

(ii) The fact that the images can be accompanied by text allows to easily identify those images that appear together with specific words; and

(iii) By considering public pictures posted by hundreds of

users, these users are not aware that they are being part of a psychological experiment.

Instagram images are of high interest because there are no boundaries in the kind of pictures that users use to communicate with their followers. Thus, different kinds of users will post different kinds of pictures, as proven by Hu et al. [20]. In their work, they show that the pictures posted in Instagram can be classified into eight main categories, and the users can be divided in five different groups, depending on what kind of pictures they post. These eight main picture categories are: friends, food, gadget, captioned pictures, pets, activity, selfies, and fashion. The difference on the type of images posted can be influenced by the city of the users, [22] or their age [23].

In order to determine the *MindPics* set of images, we used the words from Table I to query images. For each personality trait, 22 words were used, 11 for each component, and about 1,100 images were selected for each word. The total number of *MindPics* images used for training the personality model is 121, 000. Because we used the same number of words per trait and the same number of images per word, the number of training images results balanced within the 5 personality traits and also within the used tags, thus each trait being trained from around 24, 000 images.

In Figure 2 we show some 10 random samples of each personality trait obtained with the procedure described above. As it can be seen, despite the huge intra-class and inter-class variabilities of the images associated to each of the personality traits, we can show that there is consistence between the images of the same class, whose hashtags are related to the words suggested by [4].

B. Building the Personality Model

Once defined the procedure to determine which set of images S is most related to each personality trait, we next describe how we can model this relationship between Mind-Pics and personality.

In this work we have used a neural network 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, i.e. 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 on 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:

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

Where N is the number of layers in the model, and each computation layer corresponds to a non-linear function f with its own parameters θ_n found by empirical risk minimization [24].

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 use a CNN model [18,25], 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 score for each of the personality traits, we propose an all-in-one model: we propose to use 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 is independent of the others and consists on a binary classifier like the described before, which has its own loss function.

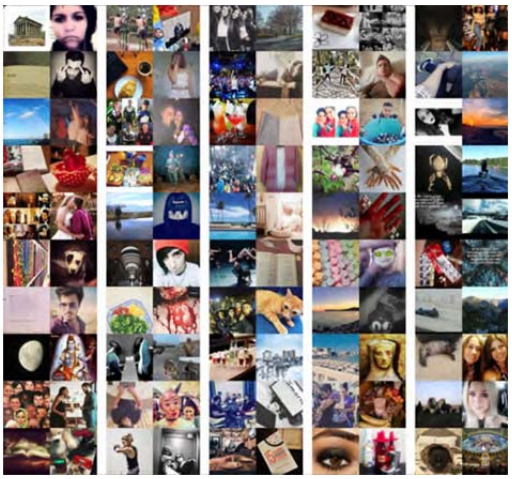

Fig. 2. Database Samples**.** Each pair of columns contains 20 *MindPics* related to each personality trait, from left to right: openness, conscientiousness, extraversion, agreeableness, and neuroticism.

The loss used for each of the 5 independent binary classifiers is the Multinomial Logistic Loss from Equation 4 as in the single-classifier model, with one minor modification. For the multi-classifier setup, we must consider that we only want to backpropagate errors made by the classifier responsible for the ground truth label. For instance, if the true label of an image is "Openness" the only error backpropagated should be the error produced by the "Openness" classifier.

IV. Experiments

In this section we describe the different models that we have tested to infer personality traits from images and explain the details of the training process. We finish the section by exposing the quantitative results of our approach and with a qualitative analysis of the model trained for personality inference.

TABLE III: Classification Accuracies (%) for each Personality Trait Based on the MindPics Dataset

Model	O	C	E	A	N	Average
AlexNet (random)	62.3	63.5	63.6	62.1	65.4	63.4
AlexNet (pre-trained)	66.9	69.2	73.6	67.8	69.4	69.4
ResNet (pre-trained)	**69.8**	**72.4**	**77.7**	**69.8**	**69.6**	**71.9**

A. CNN Models

There are some common CNN architectures [19], [26]–[28] that are well established for computer vision tasks. In our experiments, we have used two of these CNN models.

The first model (AlexNet) [19]. This architecture extracts

features from 227 227 images with five consecutive convolutional layers and predicts with three stacked Fully-Connected layers.

The other CNN model is a Residual Network presented in [28] (ResNet). Differently form AlexNet, ResNet uses residual connections between the network layers, allowing for easier optimization, and thus allowing to increase the network depth.

In the experiments we use a ResNet of 50 layers, whereas the Alexnet model consists of 8 layers. For both models we replace the last layer to contain the required amount of outputs.

B. Fine-Tuning

We test two different options for initializing the network's weights: (i) randomly sampling form a Gaussian distribution, and (ii) fine-tuning, which consists in 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 [29]. The fine-tuning approach has been proved [30] 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 these features to learn a classifier in a smaller dataset.

C. Training Setup

We trained the models with Caffe [31] on a NVIDIA GTX 770 with 4GB of memory. Models were optimized with SGD with momentum of 0.9 and batch size of 128 for *AlexNet* and *ResNet* respectively. The learning rate is 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 apply horizontal mirroring to the images and crop a random patch of 224x224 pixels of the original 256x256 images. The only pre-processing of the images is the subtraction of the training set mean. A random split of the *MindPics* dataset is used to divide the images in non-overlapping training and testing sets, 80% of the images are used for training and 20% for testing.

D. Quantitative Results

In Table III the accuracies on personality recognition for each trait obtained by the different models and configurations we tested are shown. The first row shows the performance of the AlexNet model trained from scratch with random weights initialization.

The rest of the results show that pre-training the network in a larger dataset in order to learn to extract better features increases performance in all the personality traits, increasing the average accuracy by 6%. It can also be seen that the all-in-one model performs better in this scenario too, both for all architectures.

The all-in-one configuration learns better features because the personality classifiers share the feature extraction layers and each classifier contributes to the learning, whereas a model that only has one classifier does not see as many different images, so it learns worse features. However, when pre-training on the *Imagenet*, both networks start already with good feature extractors, so this effect is not as important as when the networks are trained from scratch. Besides the increase in accuracy performance, the all-in-one network is also much more efficient, because it shares most of the image processing steps for all traits, thus reduces the amount of computation by a factor of five.

Finally, the results also show that the *ResNet* network is significantly better than the shallower network *AlexNet*, achieving up to 2.5% more in average accuracy and consistently increasing the performance for all traits. This increase in performance can be explained from the increase in depth of the model, which allow the models to learn better representations.

E. Qualitative Results

To get a better insight of what the new deep features are detecting, we visualize and analyze the images that maximally activate the output of the network for each personality trait. Namely, in order to know which images better represent each personality trait, we find those pictures that maximally activate a specific output of our model. Similarly to [32], we feed all the images to our model, inspect the activation values of a specific neuron, and look for the images that produce these maximum activations.

In our case, we inspect the output units associated to each personality trait. In Figures 3-7 the most representative *MindPics* for each personality trait are shown.

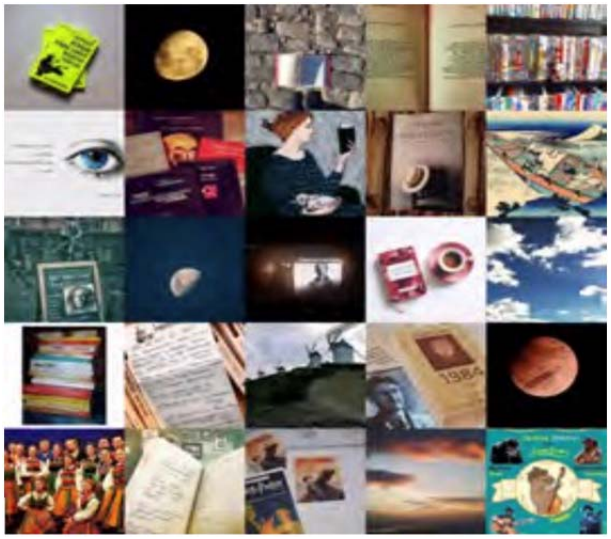

Fig. 3. *MindPics* that maximally activate the Openness trait. One can see pictures of books, moons and sky.

Fig. 4. For the Conscientiousness trait, most of the images that are maximally activated correspond to food, especially healthy one.

Fig. 5. The Extraversion trait is mostly activated by pictures of a lot of people, and scenery of parties and night life.

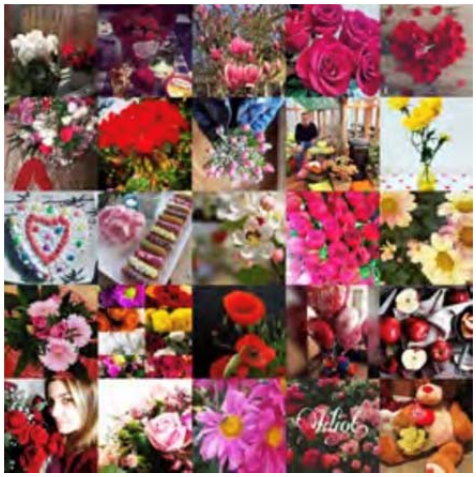

Fig. 6. In Agreeableness, the MindPics correspond to mostly flower pictures which are mostly activated.

Fig. 7. For the Neuroticism trait we observe that the score is maximally activated by pets, like cats and dogs.

V. Discussion

Social media is the product of the expression of the users by means of text, images, and speech. This is a great opportunity for companies and researchers not only to know about the content of the images that are being shared, but to know about the users that create and interact with those images themselves.

One interesting trait to learn from the users is personality. In fact, different methods have already been proposed to extract the users' personality from social media text and favorited images. This indicates that personality might be an underlying factor in the distribution of the users' data. In this work, we made a step towards confirming this hypothesis, directly focusing on the authors of the pictures and showing that personality remains invariant when moving from the text to the image domain.

In particular, we showed that given the images uploaded by those users who used words correlated to the different personality traits [4], it is possible to retrieve their personality from the images. Thus, images and text are correlated, which can be explained if both depend on a third variable, which is personality. In this study, we do not recover the full spectrum of personality traits of one user, but we infer a specific personality trait that a single image conveys.

The underlying hypothesis is that when a user posts a picture in a social network, the picture is not expressing everything the author has in mind, only the specific message intended by the author. In the same way, a picture does not describe the whole personality of the user, but a portion of it. So, in order to get 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 of their personalities.

VI. Conclusion

In this work we focus on the modeling the OCEAN personality traits by applying deep learning techniques to find visual patterns. A dataset which consists of images shared in Instagram has been built, based on those key words most correlated to a particular personality trait, as already identified in the cyberpsychology literature. From this procedure, different visual patterns of personality traits emerge, as described next. Generally, there are many images of people posing, sharing, laughing in the Extraversion class. Moreover, the network has tendency to label black and white images as Neuroticism with high confidence ratios. The network is able to learn it by itself that when there is a group of friends in an image, the network tends to classify it as Extraversion. On the other hand, if there is a pet and a person, the network generally classifies it as Agreeableness. Another interesting case is when the network classifies images of eating as Conscientiousness. Lastly, the network tends to classify an image of sky and nature as Openness.

Unfortunately, annotation using social media tags is a problem, since the downloaded dataset has not been cleaned manually. The cleaning for larger dataset is quite difficult, because checking every single image in thousands of images is not feasible. For the future work, an extra class will be added to the network for irrelevant samples. This class will stand for images which do not convey any emotion, since not all images convey feelings. In that way, the biggest problem of the project which is irrelevance caused by images could be alleviated.

Acknowledgment

Authors would like to specially thank Ms. Daniela Rochelle Kent, Mr. Vacit Oguz Yazici, and Mr. Alvaro Granados Villodre for their invaluable help with the ontology of words and the images used in the experiments. We also

gratefully acknowledge the support of NVIDIA Corporation with the donation of a Tesla K40 GPU and a GTX TITAN GPU, used for this research.

REFERENCES

[1] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *Proc. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom),* IEEE, 2011, pp. 149–156.

[2] F. Iacobelli, A. J. Gill, S. Nowson, and J. Oberlander, "Large scale personality classification of bloggers," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 568–577.

[3] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *Proc. 21st ACM international confer- ence on Multimedia*. ACM, 2013, pp. 213–222.

[4] T. Yarkoni, "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers," *Journal of Research in Personality*, vol. 44, no. 3, pp. 363–373, 2010.

[5] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3156–3164.

[6] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3128–3137.

[7] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: Lessons learned from the 2015 mscoco image captioning challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[8] J. M. Digman, "Personality structure: Emergence of the five-factor model," *Annual Review of Psychology*, vol. 41, no. 1, pp. 417–440, 1990.

[9] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: a meta-analysis," *Personnel Psychology*, vol. 44, no. 1, pp. 1–26, 1991.

[10] L. R. Goldberg, "An alternative" description of personality": the big-five factor structure," *Journal of Personality and Social Psychology*, vol. 59, no. 6, p. 1216, 1990.

[11] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, no. 2, pp. 175–215, 1992.

[12] D. P. Schmitt, J. Allik, R. R. McCrae, and V. Benet-Mart´ınez, "The geographic distribution of big five personality traits: Patterns and profiles of human self-description across 56 nations," *Journal of Cross-cultural Psychology*, vol. 38, no. 2, pp. 173–212, 2007.

[13] C. Segalin, D. S. Cheng, and M. Cristani, "Social profiling through image understanding: Personality inference using convolutional neural networks," *Computer Vision and Image Understanding*, 2016.

[14] J. Oberlander and S. Nowson, "Whose thumb is it anyway? classifying author personality from weblog text," in *Proc. COLING/ACL on Main Conference Poster Sessions*, Association for Computational Linguistics, 2006, pp. 627–634.

[15] S. C. Guntuku, S. Roy, and W. Lin, "Personality modeling-based image recommendation." in *MMM (2)*, 2015, pp. 171–182.

[16] B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[17] C. Segalin, A. Perina, M. Cristani, and A. Vinciarelli, "The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits," *IEEE Transactions on Affective Computing*, 2016.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Infor- mation Processing Systems*, 2012, pp. 1097–1105.

[19] K. Ginsberg, "Instabranding: Shaping the personalities of the top food brands on instagram," *Elon Journal of Undergraduate Research in Communications*, vol. 6, no. 1, 2015.

[20] Y. Hu, L. Manikonda, S. Kambhampati *et al.*, "What we instagram: A first analysis of instagram photo content and user types." in *ICWSM*, 2014.

[21] F. Souza, D. de Las Casas, V. Flores, S. Youn, M. Cha, D. Quercia, and V. Almeida, "Dawn of the selfie era: The whos, wheres, and hows of selfies on instagram," in *Proc. 2015 ACM on conference on online social networks*. ACM, 2015, pp. 221–231.

[22] N. Hochman and R. Schwartz, "Visualizing instagram: Tracing cultural visual rhythms," in *Proc. workshop on Social Media Visualization (SocMedVis) in Conjunction with the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM–12)*, 2012, pp. 6–9.

[23] J. Y. Jang, K. Han, P. C. Shih, and D. Lee, "Generation like: comparative characteristics in instagram," in *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 4039–4042.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[30] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.

[31] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

Pau Rodriguez received the MSc degree in artificial intelligence from KU Leuven in 2015. He is currently a Ph.D. student at the Image Social Evaluation (ISE) Lab of the Computer Vision Center and the Universitat Autònoma de Barcelona, Catalonia Spain. His research interests focus on machine learning, pattern recognition, and computer vision.

Jordi Gonzàlez received the PhD degree in Computer Engineering from Universitat Autònoma de Barcelona (UAB) in 2004. He is an associate professor in computer science at the Computer Science Department, UAB. He is also a research fellow at the Computer Vision Center, where he has co-founded three spin-offs (Cloud Size Services, Visual Tagging, Care Respite) and the Image Sequence Evaluation (ISE Lab) research group. His research interests include machine learning techniques for the computational interpretation of social images, or Visual Hermeneutics.

Josep M. Gonfaus received the PhD degree in Computer Engineering from Universitat Autònoma de Barcelona (UAB) in 2012. He participated in various Pascal challenges. He co-founded a spin-off based on their research by applying computer vision and deep learning techniques for analyzing social media data (Visual Tagging). His main research interests are deep learning techniques to mimic human cognitive capabilities.

F. Xavier Roca received the Ph.D. degree in computer science from Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain, in 1990. He is an Associate Professor and the Director of the Department of Computer Science, UAB. He is also a Research Fellow with the Computer Vision Center. He has been a Principal Researcher in several projects (public and private funds). He is working in technological transfer computer vision. The topics of his research are active vision, biometrics and tracking.