Computer Vision and Image Understanding 000 (2016) 1–17

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Social profiling through image understanding: Personality inference using convolutional neural networks

Cristina Segalin[a,*], Dong Seon Cheng[b], Marco Cristani[c]

[a] Department of Computing and Mathematical Sciences, California Institute of Technology, Computational Vision Lab, 1200 E California Blvd, 91125-0001 Pasadena, CA
[b] Department of Computer Science and Engineering, Hankuk University of Foreign Studies, Yongin, South Korea
[c] Department of Computer Science, University of Verona, Verona, Italy

A R T I C L E I N F O

Article history:
Received 28 December 2015
Revised 30 September 2016
Accepted 19 October 2016
Available online xxx

Keywords:
Image understanding
Social signal processing
Convolutional neural networks
Computational aesthetics
Personality computing

A B S T R A C T

The role of images in the last ten years has changed radically due to the advent of social networks: from media objects mainly used to communicate visual information, images have become *personal*, associated with the people that create or interact with them (for example, giving a “like”). Therefore, in the same way that a post reveals something of its author, so now the images associated to a person may embed some of her individual characteristics, such as her personality traits. In this paper, we explore this new level of image understanding with the ultimate goal of relating a set of image preferences to personality traits by using a deep learning framework. In particular, our problem focuses on inferring both self-assessed (how the personality traits of a person can be guessed from her preferred image) and attributed traits (what impressions in terms of personality traits these images trigger in unacquainted people), learning a sort of wisdom of the crowds. Our characterization of each image is locked within the layers of a CNN, allowing us to discover more entangled attributes (aesthetic patterns and semantic information) and to better generalize the patterns that identify a trait. The experimental results show that the proposed method outperforms state-of-the-art results and captures what visually characterizes a certain trait: using a deconvolution strategy we found a clear distinction of features, patterns and content between low and high values in a given trait.

1. Introduction

Two directions have shaped the image understanding field of the last 30 years (Liu et al., 2007): the first is the one of the low-level processing, where basic information is extracted from the pixel values in the form of color histograms, frequency responses etc., and used to create a representation in a vectorial space, where tasks of clustering or classification can be carried out (Carson et al., 1999; Vailaya et al., 2001). In the second direction, the semantic content of the image is extracted by means of segmentation, classification and detection approaches, and used for tasks such as content-based indexing and retrieval (Li et al., 2010; Smeulders et al., 2000).

The advent of Internet, the capability of dealing with big data, and the diffusion of social media, gave rise to a third way of dealing with images (Jin et al., 2011; Vinciarelli and Pentland, 2015); specifically, images started being associated with *people*: in facts, images are now digital objects that could be easily uploaded by a certain user into social platforms such as Facebook, Flickr, and the like. Images can be also tagged as “preferred”, highlighting those shots that naturally meet expectations of one in terms of aesthetical preferences and/or semantic content.

Both of these activities (uploading and tagging pictures) indicate a substantial revolution in how images are used: from means to represent visual aspects of reality, where the ownership of the photo is neglected, they have become *personal messages*, from the sender (the subject which uploads the photos into a social network, or that selects some shots as favorite) to his receiver(s) (the user of the social network that sees the uploaded or the preferred pictures). In this fresh new perspective, uploading or “preferring” images will communicate something, that is, personal messages as the kind of subjects that one may like (cars, landscapes, people) or the life experiences one is going through. But images communicate more than this, and this fact does represent a true revolution in the image understanding field, with a new layer of image interpretation which has started to be unveiled; to explain this new perspective, the sender/receiver communication perspective discussed above becomes invaluable.

* Corresponding author.
E-mail addresses: cristina.segalin@univr.it (C. Segalin), cheng_ds@hufs.ac.kr (D.S. Cheng), marco.cristani@univr.it (M. Cristani).

http://dx.doi.org/10.1016/j.cviu.2016.10.013
1077-3142/

In dyadic face-to-face communications, people share their opinions, experiences and impressions of life by using explicit verbal signals (that is, spoken sentences) and non-verbal signals (for example, by how they deliver the sentences, or by assuming bodily expressions) (Vinciarelli and Mohammadi, 2014; Vinciarelli and Pentland, 2015). Many social psychology studies highlight the fundamental importance of both aspects, the verbal content and the non-verbal signals, for the successful exchange of messages. This two-body communication paradigm is modeled by the Brunswick lens, in the field of social psychology (Brunswik, 1956).

Very recently, the Brunswick lens model has been customized for this new kind of communication by images: in this new setting, personality traits have been considered as the social signals sent with the uploaded images, and whose inference is one of the most intriguing challenges. In this respect, the works of Cristani et al. (2013); Segalin et al. (2016) focused on inferring with a regressor the real personality traits of the sender (collected by self-assessed tests), but also those traits that unacquainted people (the *assessors*) associate with the sender by looking at her images. In particular, Segalin et al. (2016) showed that the assessors' evaluations were 1) consistently similar, 2) in partial disagreement with the self-assessed evaluations, and 3) more easily predicted by machine learning techniques. In other words, the act of sharing images online may evoke a common psychological response in the receiving crowd, and this can be reasonably predicted by automatic approaches. Thus, it is possible to build a *wisdom of the crowds* model of personality profiles from collections of images, based on the impressions these may generate on a general hypothetical audience.

A limitation of the approach in Segalin et al. (2016) is that the features used to describe the images are taken from the computational aesthetics (CA) literature; in practice, CA often focuses on designing features that explain how a particular image has been captured, discarding the content of the images. In addition, given the wide spectrum of subjects appearing in database images, standard object recognition and feature extraction techniques might not be sufficient to capture significant dependencies between the pictures and the personality traits of their owner. This leads to the development of more advanced techniques such as feature learning, carried out in this paper by convolutional neural networks.

Computer vision with convolutional neural networks (CNNs) has received much attention in recent years, as it is well suited for processing large amounts of data and providing outstanding performances in classical problems like object (Krizhevsky et al., 2012) and image style (Karayev et al., 2013) recognition. In fact, our approach fine-tunes CNNs pre-trained for image classification with the intention of co-opting their effective representational power to indirectly capture the aesthetic attributes of photographs, with the ultimate goal of predicting the personality traits associated with them. This allows us to discover more entangled attributes and to better generalize the patterns that identify a trait. In practice, whereas CA features are explicitly crafted to reveal information about the style of an image, remaining agnostic w.r.t. the content of the image, CNNs exhibit no such limitation, capturing both the aesthetic patterns in the pictures and their content, unveiling semantic information (for example, capturing possible recurrent objects preferred by a user).

Experiments have been focused on the *PsychoFlickr* corpus (Segalin et al., 2016): the dataset provides 200 "favored" images from 300 Flickr users for a total of 60,000 images. Additionally, the personality profile of each user is described in terms of the Big Five traits (Rammstedt and John, 2007) extensively used in psychology: *Openness to experience* (O), *Conscientiousness* (C), *Extraversion* (E), *Agreeableness* (A) and *Neuroticism* (N). This information is collected both through a self-assessment questionnaire and an independent group of 12 assessors, rating the image sets of each user. This allows the corpus to supply two different evaluation criteria for the same data.

The experimental results show that the proposed method sufficiently captures what characterizes a certain trait: on a quantitative level, it performs around 10% better on attributed traits than on self-assessed ones, with a best accuracy of 68% on attributed Neuroticism; on a qualitative level, ranking the test images by confidence shows a clear distinction of features, patterns and content between low and high values in a given trait. These results also outperform (Segalin et al., 2016) when suitably re-casted from regression to classification. Finally, we also introduce an online application demo that uses our trained classifiers to predict personality traits given a proposed set of pictures liked by a subject.

In the following sections, we first describe some related work in computer vision and computational aesthetics; we then introduce our approach based on processing the *PsychoFlickr* corpus using *convolutional neural networks*, followed by a section discussing the results. Finally, we briefly present our demo and provide some concluding remarks.

2. Related work

The idea that aesthetic values are connected to features goes back at least to Birkoff in the 1930s (Birkhoff, 1933). Hoenig (Hoenig, 2005) in 2005 comprehensively defined computational aesthetics (CA) as a field of study with many emphasis on three important factors: computational methods, the human aesthetic point of view and the need to focus on objective approaches. CA is an inter-disciplinary area at the crossroad between computer vision and pattern recognition (CVPR), psychology, visual art aesthetics and neuroscience (Joshi et al., 2011). Learning the aesthetic appeal of images from the wisdom of the crowds has been used in many studies by using explicit aesthetic scores (Dhar et al., 2011; Murray et al., 2012), or relying directly on textual tags and likes (called "*favs*" on Flickr) (Bauckhage and Kersting, 2013). On the other hand, psychology and neuroscience have investigated the influence of individual characteristics on aesthetic preferences for 70 years (Joshi et al., 2011). This applies to style (Furnham and Walker 2001) and, to a lesser extent, content of paintings (Rawlings et al., 1998). The main result of these investigations is the identification of correlations between artistic preferences and personality (Furnham and Walker 2001), where the latter is typically described in terms of the Big Five traits (Rammstedt and John, 2007). For example, considering content-based aspects of paintings, a link has been detected between Openness and preference for pictures with few elements, as well as for polygons rated as "complex" and "meaningful" (Rawlings et al., 1998). Datta in 2006 introduced a computational approach for the investigation of features that underpin aesthetic value in artistic and photographic images (Datta et al., 2006). Many subsequent studies have applied machine learning to photographic digital images and/or art images with the goal of predicting the aesthetic rating of these images from features engineered to measure image properties such as colorfulness, brightness and texture (for example, Bauckhage and Kersting (2013); Campbell et al. (2015); Ciesielski et al. (2013); Datta et al. (2006); Dhar et al. (2011); Furnham and Avison (1997); Galanter (2012); Gong); Lu et al. (2014); Machado and Cardoso (2010)). The original features of Datta have been adopted for the analysis of artworks in Ciesielski et al. (2013), while an extension of the original features has been proposed in Datta (2009); Ke et al. (2006). As machine learning extends learning from engineered features sets towards a new paradigm of representational learning and features discovery, computational aesthetics has begun to exploit these capabilities also. Lu et al. (2014) presented a method based on multi-column deep convolutional networks for predicting the aesthetic rating of photographic images. When it comes to personality and images, the literature is scarce as it is a very new research challenge. Most of them took into account other social networks platforms as Facebook and Twitter (Golbeck et al., 2011; Gosling et al., 2007; Hughes et al., 2012; Kosinski et al., 2013; Quercia et al., 2011; 2012), and used other information publicly available on the user profile.

The first works experimenting on images and personality traits have been Cristani et al. (2013); Segalin et al. (2016), where the authors predicted both self and attributed traits from the favorite images of a pool of 300 users, in the *PsychoFlickr* corpus. In this case they used a hybrid approach where generative models, used as latent representations of the features extracted from the images, are built and then passed to discriminative classifier to predict the personality trait of each user of the pool. A similar direction is followed by deep learning, where new model-based representations for input data are automatically discovered, in the form of salient features that are more powerful than any features set of human design. Indeed, in this work we investigate the possible power of this new representation.

3. Our approach

The groundbreaking success of the Convolutional Neural Networks (CNNs) in the ILSVRC challenges (Krizhevsky et al., 2012) have clearly demonstrated the aptitude of these classifiers at deconstructing the elements and features contained within photographs. Most importantly, they share some basic primitive components analyzed in the first works of personality inference (which essentially applied standard CA features): color, composition, textural properties, etc. More in the detail, CNNs represent a more effective way to combine such primitive components into holistic representations, aimed not only at highlighting stylistic patterns (as done for example in Karayev et al. (2013), and in general by all the computational aesthetics works), but also to unveil information about the content of the image. To investigate this idea, we adopt CNNs pre-trained for image classification and fine-tune them to predict the personality traits contained in the *PsychoFlickr* dataset. In this way, we leverage the deconstructing power of networks trained on millions of images in order to learn visual representations correlated with psychological profiles. The upside of this approach is that, since pre-trained networks are learned on a large set of images, the intermediate layers capture the “semantics” of the general visual appearance in a way superior to hand-crafted features. The downside is represented by the “black box” nature of the technique: the insights discovered and codified by the network are locked within its connectivity structure, only revealed approximately by considerable analysis work (Zhou et al., 2014b). However, this black box can be effectively exploited by adding a few layers on top, tailored for our specific problem.

3.1. Psychoflickr: pictures and personality

The *PsychoFlickr* corpus contains information on 300 Flickr accounts, simply called users in the following. For each user, there are 200 random pictures taken from those he “liked”. Crucially, these are not pictures *made by that user*, in order to remove biases due to his personal photographic skills.

Furthermore, the dataset contains self-assessed and attributed personality traits for each user. The former are based on a personality questionnaire, the BFI-10 (Big Five Inventory 10) (Rammstedt and John, 2007), filled in voluntarily by each user, and the latter are based on the evaluations of 12 independent *assessors* that looked at each user’s favorite images and filled in a modified BFI-10 questionnaire.

The outcome of the BFI-10 is a five dimensional vector where each component measures the tendency of an individual with respect to the Big Five traits, namely *Openness* (tendency to be intellectually open, curious and have wide interests), *Conscientiousness* (tendency to be responsible, reliable and trustworthy), *Extraversion* (tendency to interact and spend time with others), *Agreeableness* (tendency to be kind, generous, etc.) and *Neuroticism* (tendency to experience the negative aspects of life, to be anxious, sensitive, etc.). In particular, the BFI-10 provides an integer score ranging from -4 (low tendency) to 4 (high tendency).

Table 1
Number of users selected for the binary classification problems: for each combination of trait and label (self-assessed or attributed), we report the size of the low and high classes.

	Self-assessed traits					Attributed traits				
	O	C	E	A	N	O	C	E	A	N
Low	113	122	75	110	104	81	77	79	82	80
High	119	121	76	125	111	75	80	75	75	77

Therefore, for each user u, we have a label ${}^{(self)}y_p^u$ (with $p = O, C, E, A, N$), which is called the *self-assessed* personality signature, and a label ${}^{(attr)}y_p^u$, called *attributed* personality signature, obtained by averaging the 12 ratings from the assessors. Although biased towards the portrait of themselves people try to convey, self-assessed traits are traditionally considered to be the true personality of an individual (Rammstedt and John, 2007), while attributed traits show how the subjects are *perceived* by others. Fig. 1 shows the distributions of values for both self and attributed traits.

3.2. Personality classification by fine-tuning convolutional neural networks

In our approach, we decide to simplify the task of predicting the personality traits in the *PsychoFlickr* corpus into five distinct binary classification problems, one for each trait. According to the application of Factor Analysis to behavioral evidence (Costa et al., 1992), the Big-Five traits are independent. The range of values for each trait is partitioned into three sections: low set, for values below the first quartile; high set, for values above the third quartile, and middle set (see Fig. 1). We select only the users with values included in the low and high sets for our binary classification problems in order to get a greater separation between the two classes. From Table 1, where we report the number of users included in each class, we note that the selected subsets are roughly balanced, with self-assessed classes having more than 100 users (except for Extraversion) and attributed classes having around 80 users each. Considering both self-assessed and attributed traits, we thus have 10 independent datasets of images with binary labels. As protocol, we keep 75% of each dataset for training and 25% for testing.

In this work, we consider CNNs that have been trained to compete on the ImageNet 2012 challenge, like AlexNet, the eight-layer network that won in 2012 (see Krizhevsky et al. (2012) for details), and change the last layer in order to adapt it to our binary classification problems. These trained nets are ideal candidates for fine-tuning because they have been trained on a large number of images (1.2 million) and a wide number of classes (1000 object categories), providing for a very solid acquisition of representational power.

The layers in the network are conceptually split into two blocks, the early convolutional layers (with some max-pooling layers interleaved) and the later fully connected layers: the first block processes an input image in terms of visual patterns, and the second block seeks a higher-level semantic representation based on those visual patterns. By fixing the first block, we exclusively relied on the same visual patterns of the ImageNet problem, which luckily cover a very wide spectrum due to the huge number of images and classes. This allowed the network to focus just on finding the

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

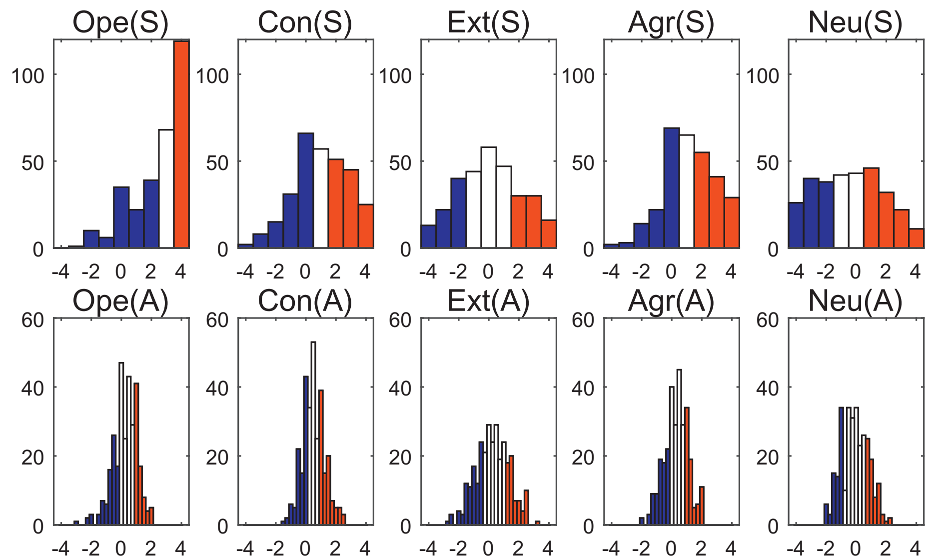


Fig. 1. The distribution of values in the self-assessed (top row) and attributed (bottom row) personality traits, with the low and high classes highlighted in blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Accuracies (Acc) on the training and test sets and average performance (AP) on the test set using AlexNet, VeryDeep-16 and Places features and linear SVMs.

		Self-assessed traits					Attributed traits				
		O	C	E	A	N	O	C	E	A	N
AlexNet	Train Acc	0.56	0.56	0.56	0.55	0.55	0.61	0.65	0.65	0.64	0.68
	Test Acc	0.53	0.54	0.53	0.53	0.52	0.60	0.65	0.64	0.63	0.67
	Test AP	0.55	0.55	0.55	0.56	0.54	0.61	0.71	0.69	0.65	0.71
VeryDeep-16	Train Acc	0.56	0.56	0.56	0.56	0.56	0.63	0.68	0.66	0.65	0.70
	Test Acc	0.53	0.55	0.54	0.54	0.53	0.62	0.65	0.64	0.64	0.68
	Test AP	0.55	0.55	0.55	0.56	0.54	0.62	0.71	0.69	0.65	0.71
Places	Train Acc	0.56	0.58	0.55	0.56	0.56	0.64	0.67	0.66	0.66	0.70
	Test Acc	0.53	0.54	0.54	0.53	0.53	0.62	0.66	0.65	0.63	0.67
	Test AP	0.54	0.54	0.54	0.57	0.54	0.60	0.70	0.68	0.65	0.70

right mix of visual patterns to characterize the psychological traits. One possible drawback is an understandable bias towards objects rather than styles within the images since the ImageNet categories are object-based.

Moreover, we controlled the overfitting tendency by raising the value of the weight decay parameter, responsible for regularizing the weights and reducing the model's training error (Krizhevsky et al., 2012), and lowering the learning rate so as to let the weights update very gradually. In the experimental section, we describe all the parameters setups.

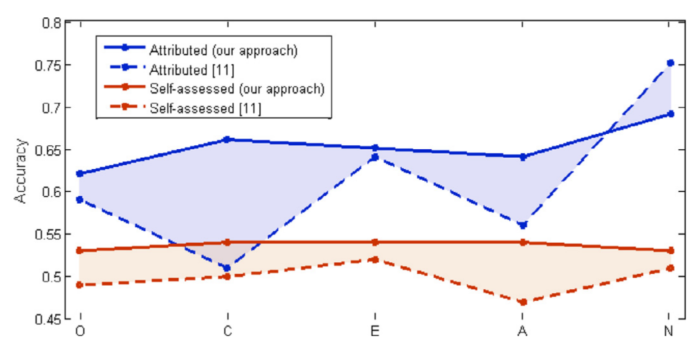


Fig. 2. Comparison of the classification accuracies between Segalin et al. (2016) and our AlexNet results.

4. Experiments and results

In this section, we first describe a set of baseline experiments where CNNs are used as feature extractors and classification is performed with linear SVMs; we then apply our approach of fine-tuning pre-trained nets and compare these results against the literature and the baseline ones; in addition, we examine a few experiments on the original regression problem in the *PsychoFlickr* dataset; finally, we analyze the attributes learned by our models. All experiments were performed on a linux pc with an NVIDIA graphical card to exploit GPU speed-ups, employing the open source software implementations of Caffe (Jia et al., 2014) and MatConvNet (Vedaldi and Lenc, 2015).

4.1. Baseline experiments with CNNs and linear SVMs

To contrast and compare against our approach, we describe here a set of baseline experiments with traditional setups: features are extracted from the dataset images and linear SVMs are used to solve the classification problems. Results can vary greatly on the choice of feature extractors and the most relevant choice in our situation is to use the CNNs themselves.

In particular, it has been shown that activation vectors collected before the last layer of a net, which acts not unlike a linear classifier, are the most informative as features. Thus, we proceed in this way: images are isotropically rescaled to have the small-

Table 3
Classification results in terms of accuracies (Acc) and average performance (AP) by fine-tuning AlexNet and VeryDeep-16 according to our approach.

		Self-assessed traits					Attributed traits				
		O	C	E	A	N	O	C	E	A	N
AlexNet	Train Acc	0.57	0.59	0.60	0.57	0.55	0.73	0.81	0.76	0.76	0.81
	Test Acc	0.53	0.54	0.54	0.54	0.52	0.61	0.66	0.64	0.64	0.68
	Test AP	0.54	0.55	0.55	0.55	0.54	0.62	0.71	0.69	0.65	0.71
VeryDeep-16	Train Acc	0.55	0.55	0.55	0.54	0.55	0.61	0.66	0.64	0.63	0.68
	Test Acc	0.54	0.55	0.54	0.54	0.54	0.61	0.67	0.65	0.64	0.69
	Test AP	0.55	0.56	0.55	0.56	0.55	0.62	0.73	0.70	0.65	0.72

Table 4
Confusion matrices for the test set in the self (top rows) and attributed (bottom rows) traits.

Self-assessed										
	Low	High	Low	High	Low	High	Low	High	Low	High
Low	**0.30**	0.19	**0.29**	0.21	**0.27**	0.23	0.21	0.26	0.10	0.39
High	0.28	0.23	0.25	0.25	0.23	**0.27**	0.20	**0.33**	0.09	**0.42**
	O		C		E		A		N	
Low	**0.60**	0.16	0.27	0.22	**0.36**	0.15	**0.34**	0.18	0.29	0.22
High	0.22	0.26	0.11	**0.40**	0.20	0.29	0.18	0.30	0.09	**0.40**
	Low	High	Low	High	Low	High	Low	High	Low	High
Attributed										

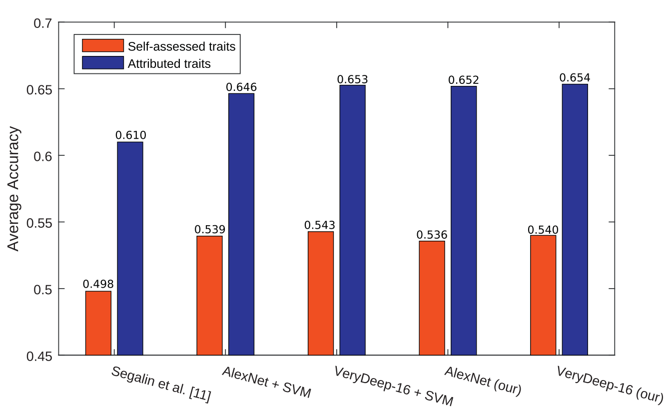


Fig. 3. Comparison of performances in terms of average accuracy across each subset of traits.

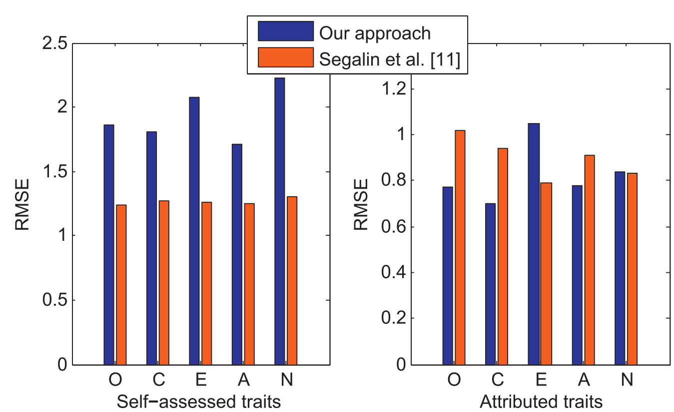


Fig. 4. Comparison of regression performances in Segalin et al. (2016) and by fine-tuning a CNN regressor base on VeryDeep-16 in terms or RMSE (lower is better).

est side at 256 pixels, then the center 227 × 227 crop is taken (this is used both in train and test mode); the feature vectors are collected after the ReLU (Rectified Linear Unit) component of the penultimate layer to reduce noise in the data, and used to train a linear SVM classifier for each of the binary problems outlined in Section 3.2. We optimize the C parameter for the SVMs by 3-fold cross-validation.

In Table 2, we report the results obtained using the 8-layers deep AlexNet architecture (Krizhevsky et al., 2012), the 16-layers deep Visual Geometry Group (VGG) architecture (Simonyan and Zisserman, 2014) (called VeryDeep-16 in the following), the 16-layers deep VGG architecture (called Places in the following) trained on the Places Database. The performances are shown in terms of train and test accuracy and test average performance (AP) of classification (the area under the precision-recall curve).

Although both AlexNet and VeryDeep-16 were trained on the same ImageNet database, winners respectively of the 2012 and 2014 challenges, the latter expresses a different philosophy, with more layers and small field convolutional filters. However, we can assume they share similar characteristics, and, in order to widen our analysis, we use Places (Zhou et al., 2014a; 2014b), which was trained on a database containing images of places to tackle the problem of scene recognition.

All three feature extractors lead to similar classification performances, with the deeper nets narrowly beating AlexNet. Regardless of the bad (on self-assessed traits) and the good (on attributed traits) accuracies, almost all training samples are used as support vectors in the SVM models, indicating a failure to capture the data effectively.

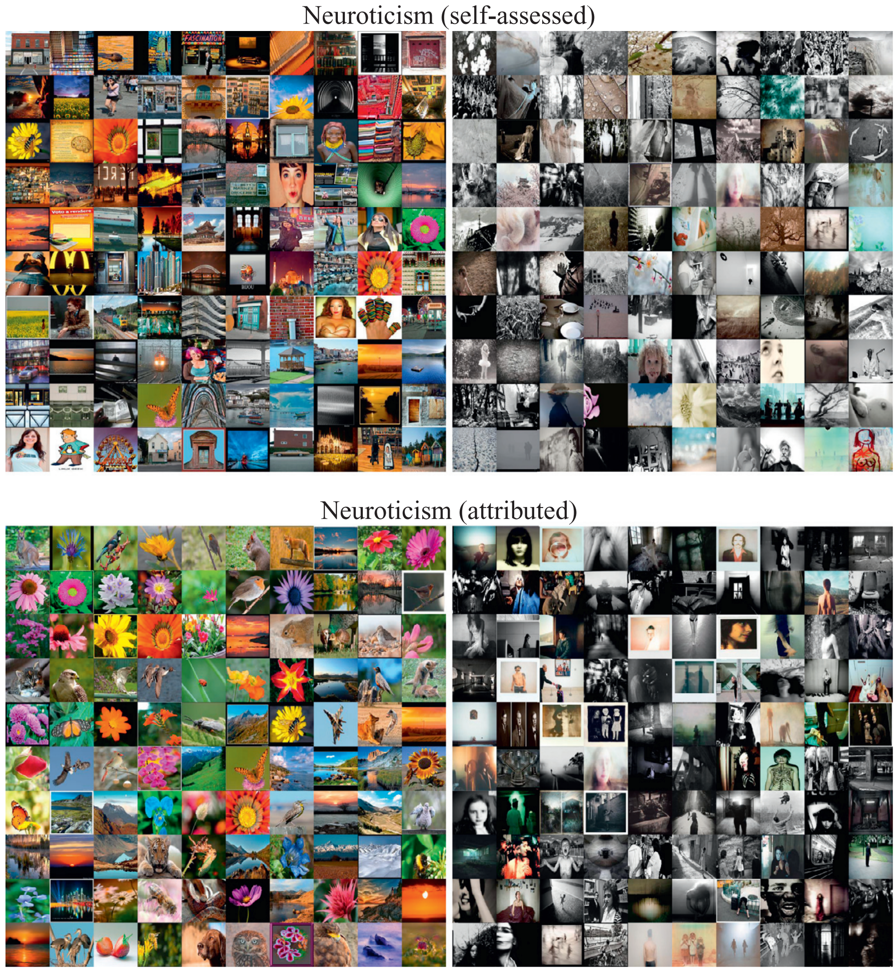


Fig. 5. Representative images in the low (left) and high (right) value classes for the Neuroticism trait.

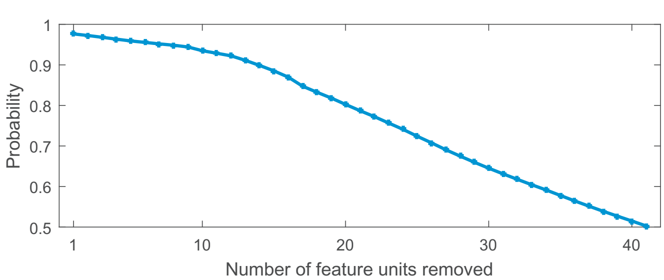


Fig. 6. Progressive drop in class probability while greedily removing feature units from the average feature map for low attributed Neuroticism: after 41 removals the probability drops below 0.5.

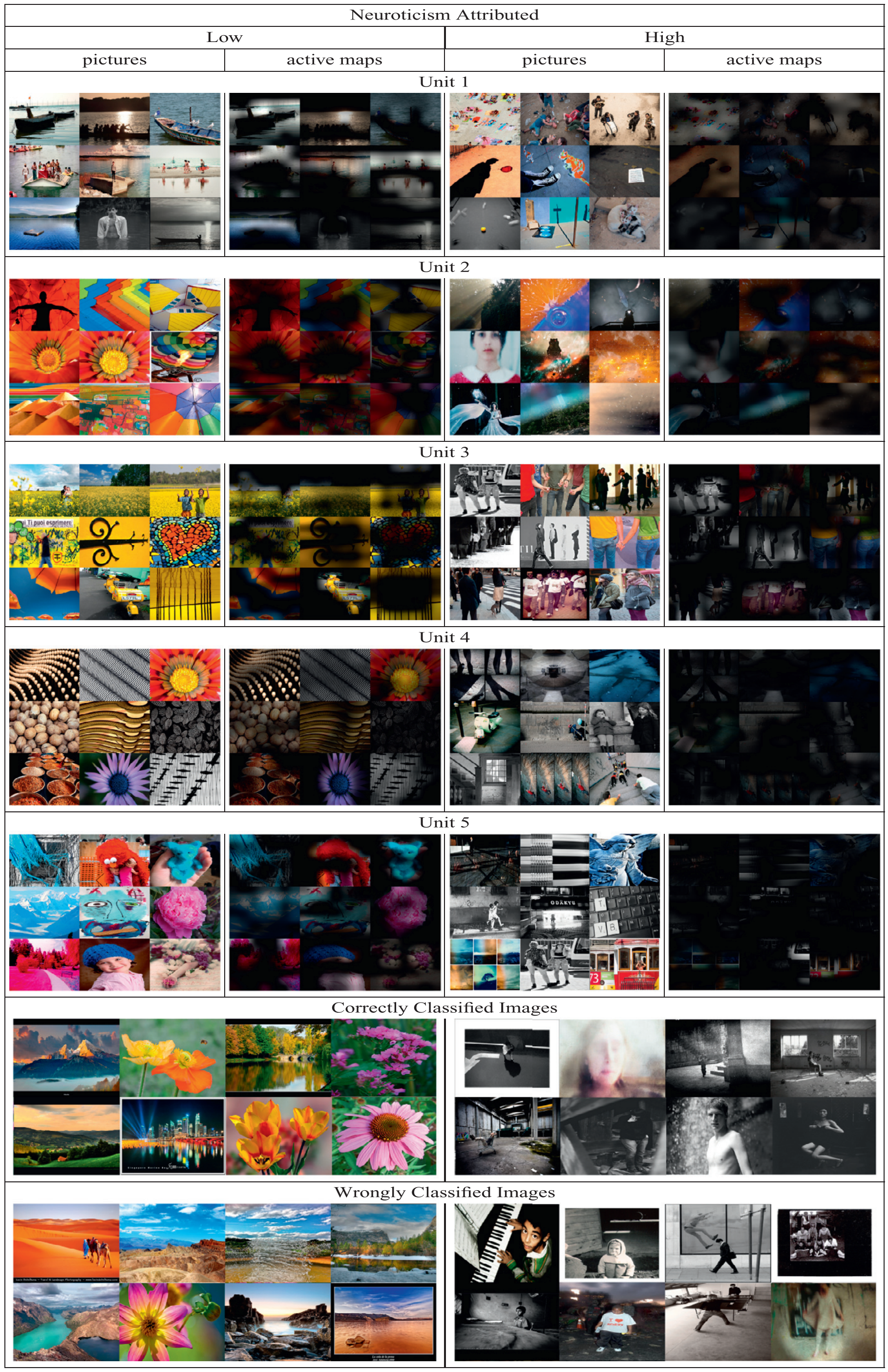
4.2. Binary classification by fine-tuning CNNs

Following the procedure described in Section 3.2, we apply our approach by fine-tuning AlexNet and VeryDeep-16 after fixing the early convolutional layers in the nets and substituting the last layer to fit our binary problem. The learning parameters are set to the following values: batch size of 50 images, momentum of 0.9, weight decay of 0.1, base learning rate of 0.0001, one step-down in learning rate to 0.00001 halfway down the 60 cycles through the training set required to complete the learning process.

Similarly to the baseline results, there is a consistent difference in performance between the self-assessed and the attributed traits. The former achieve a test accuracy roughly 10% lower than the latter, as shown in Table 3, with attributed Openness clearly harder to classify than the other attributed traits. A possible reason for the discrepancy is that the judges are obviously unacquainted with the users, and, thus, they form an impression of the users' personalities exclusively through the pictures, unaware of any ulterior motivation like personal history, inner state, education, etc. Furthermore, the consensus across the judges is statistically significant. These two factors provide a more consistent labeling that is sometimes at odd with the users' self-assessment, but this consistency of correlation between visual features and trait scores allows higher performances. In Table 4, we show the confidence matrices for the test set in the AlexNet results, with the dominant class highlighted in red: in the attributed traits, the low class is dominant for Openness, Extroversion and Agreeableness, with the high class for Conscientiousness and Neuroticism.

In Fig. 2, we compare our AlexNet results with Segalin et al. (2016), where the goal is to perform regression on the traits scores. That approach consists in a first step of dimensionality reduction for the (computational aesthetic) features characterizing the images by means of 2-dimensional counting grids (Jojic and Perina, 2011), followed by LASSO regression (Tibshirani, 1994). To obtain a fair comparison with our approach, we modify the framework of that work with our classification pipeline: we select only the users in our binary problems, and once regression is carried out by LASSO, we threshold the predicted values with the population

Table 5
Display of the 5 most significant visual attributes for attributed Neuroticism, by montages of 9 images that strongly activate the corresponding feature units, together with activation images. The bottom rows show sets of correctly and wrongly classified images.



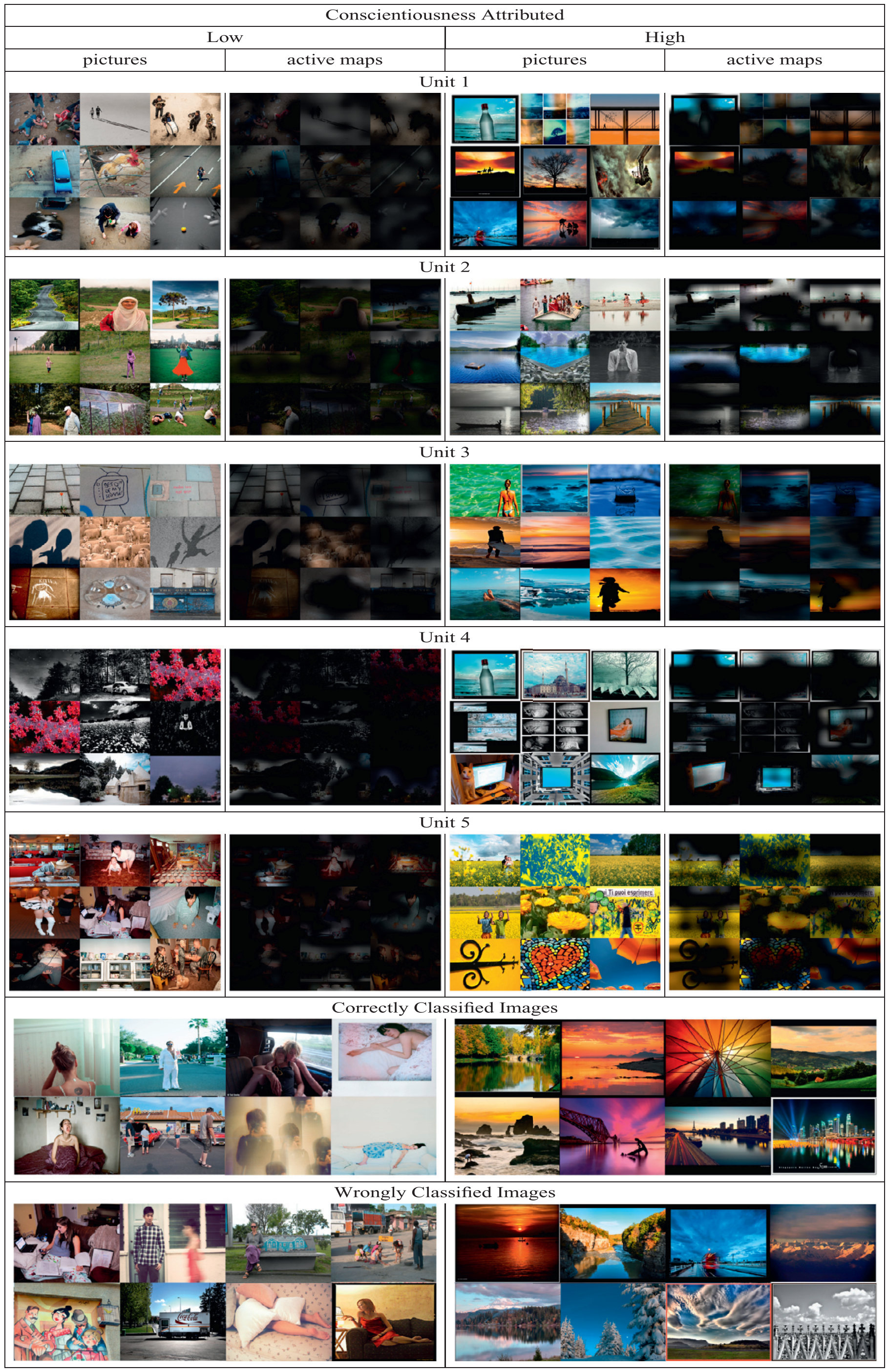
mean score for a given trait to divide predictions into low/high classes.

In Fig. 3, we compare different results in terms of average accuracy on self-assessed and attributed subsets separately. Although all our results have similar average values, within statistical fluctuations, we believe the fine-tuning approach provides additional benefits: in particular, we can analyze the details of the learned models, as discussed in the Section 4.4.

4.3. Prediction of personality traits by regression

In contrast with our approach detailed in Section 3.2 and closer to the one originally proposed in Segalin et al. (2016), here we consider predicting personality traits as a regression problem. We experimented both with training linear regressors on top of features extracted using CNNs, and with fine-tuning CNNs in regression mode. The latter approach consists in replacing the last layer

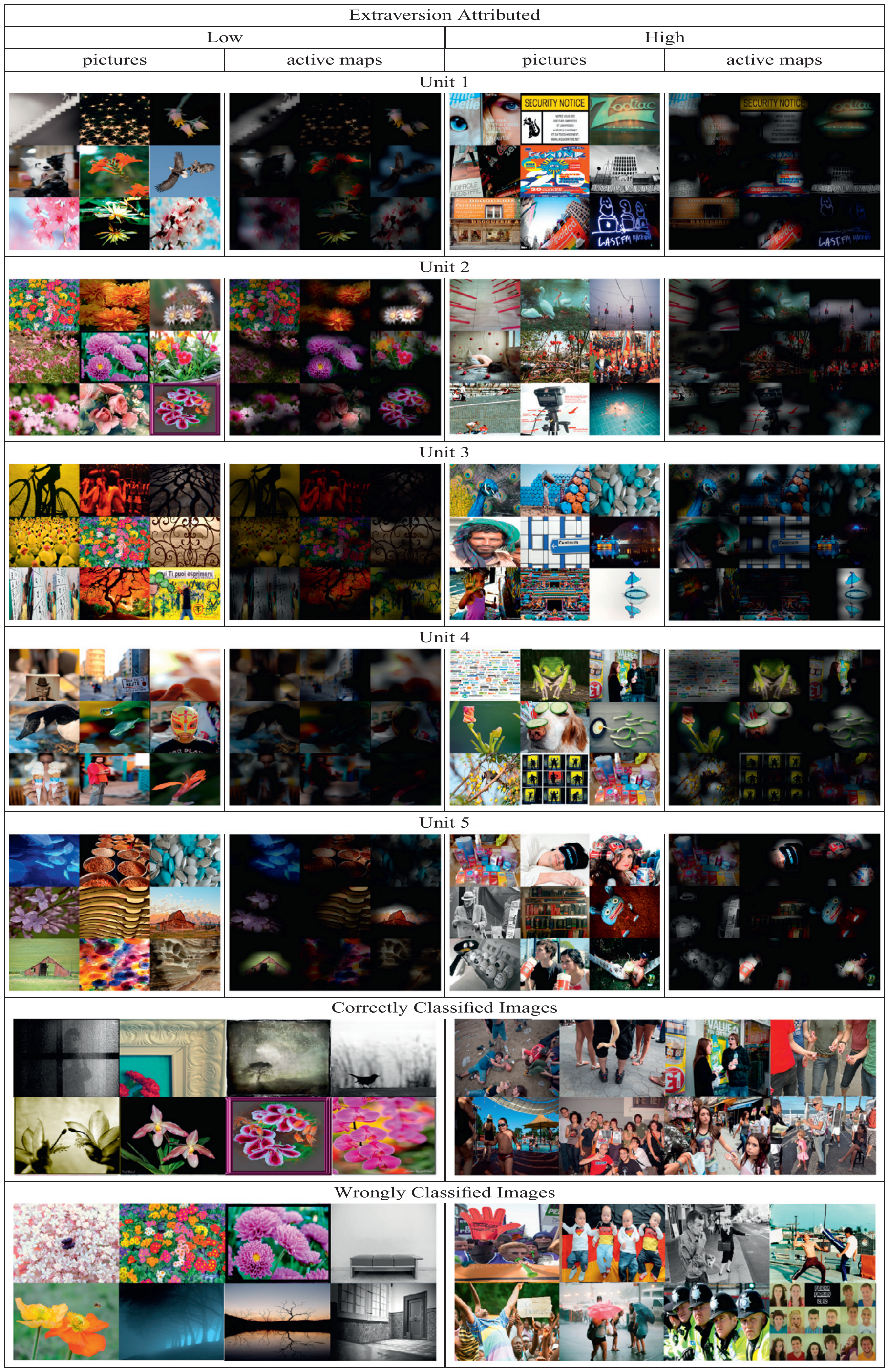
Table 6
Display of the 5 most significant visual attributes for attributed Conscientiousness, by montages of 9 images that strongly activate the corresponding feature units, together with activation images. The bottom rows show sets of correctly and wrongly classified images.



of a pre-trained net with a regression layer that minimizes a given loss function between predictions and label values, usually the Euclidean distance. After some initial setbacks, we made this setup work by adopting solutions similar to the ones in our main approach: fixing the convolutional layers and altering the learning parameters.

We considered AlexNet and VeryDeep-16 as feature extractors and base architectures for the fine-tuning as regressors. As there is no need to binarize the problem, we used the entire dataset with all the users and their original labels. All results were very similar and not too distant from the ones found by Segalin et al. (2016), with the marginally best setup obtained by fine-tuning VeryDeep-16. We show in Fig. 4a comparison between the latter and the competition in terms of root mean square error (RMSE) on the test set: we perform quite worse on the self-assessed traits, while be-

Table 7
Display of the 5 most significant visual attributes for attributed Extraversion, by montages of 9 images that strongly activate the corresponding feature units, together with activation images. The bottom rows show sets of correctly and wrongly classified images.

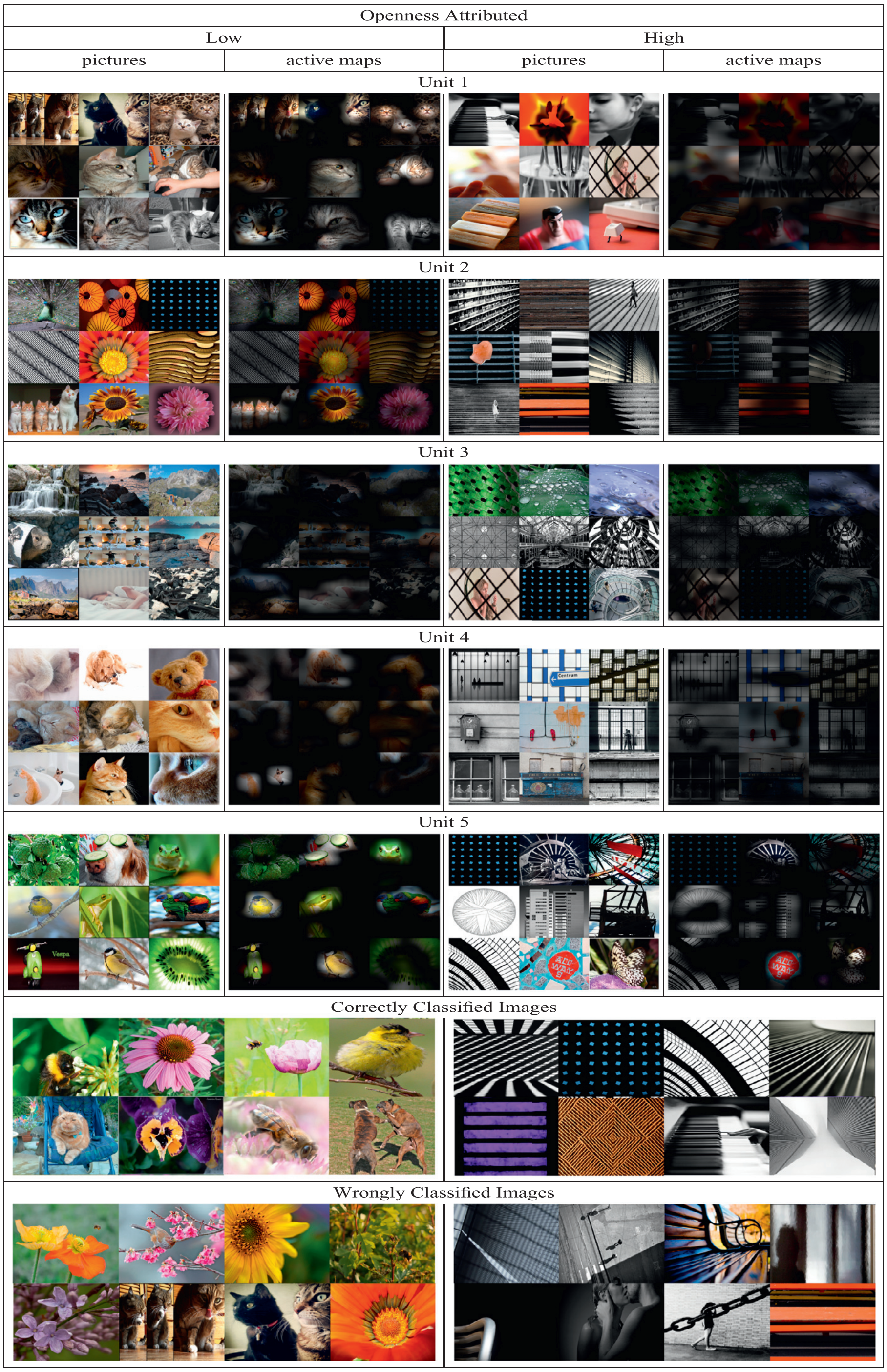


ing uneven on the attributed ones. Overall, it seems the task remains difficult to solve.

4.4. Attributes learned for each personality trait

In this section, we present three sets of analyses on the results of our approach, in order to highlight the attributes learned by the fine-tuned models for each personality trait. In particular, we focus on the VeryDeep-16 nets, as they provide marginally better results. To start, we show in Figs. 5–10 some highly representative images for the low and high class in each psychological trait. To identify these images, we first collect the output values of the softmax unit at the end of the net, and then use these values to rank correctly classified test images in decreasing order of confidence. These collections of high probability pictures show some recognizable trends and differences among the given

Table 8
Display of the 5 most significant visual attributes for attributed Openness, by montages of 9 images that strongly activate the corresponding feature units, together with activation images. The bottom rows show sets of correctly and wrongly classified images.



classes, including black-and-white versus colorful, natural versus orderly.

In addition, we identify some of the most significant features for each class and show what they represent visually. Since in our setup we keep the convolutional layers fixed, our tuned nets can be split into two distinct blocks: a) the convolutional layers extracting the features ($7 \times 7 \times 512$ feature maps, not to be confused with the feature vectors in Section 4.1), and b) the fully connected layers performing nonlinear classification of those features. Thus, understanding what the net does can be split into two distinct tasks: find what features are strongly associated with high/low classifications; and back-project features to image space to find their visual characteristics.

There are different useful protocols to find out which feature maps are significant for a certain class. Since the goal is to find some "invariant" features across images that are strongly associ-

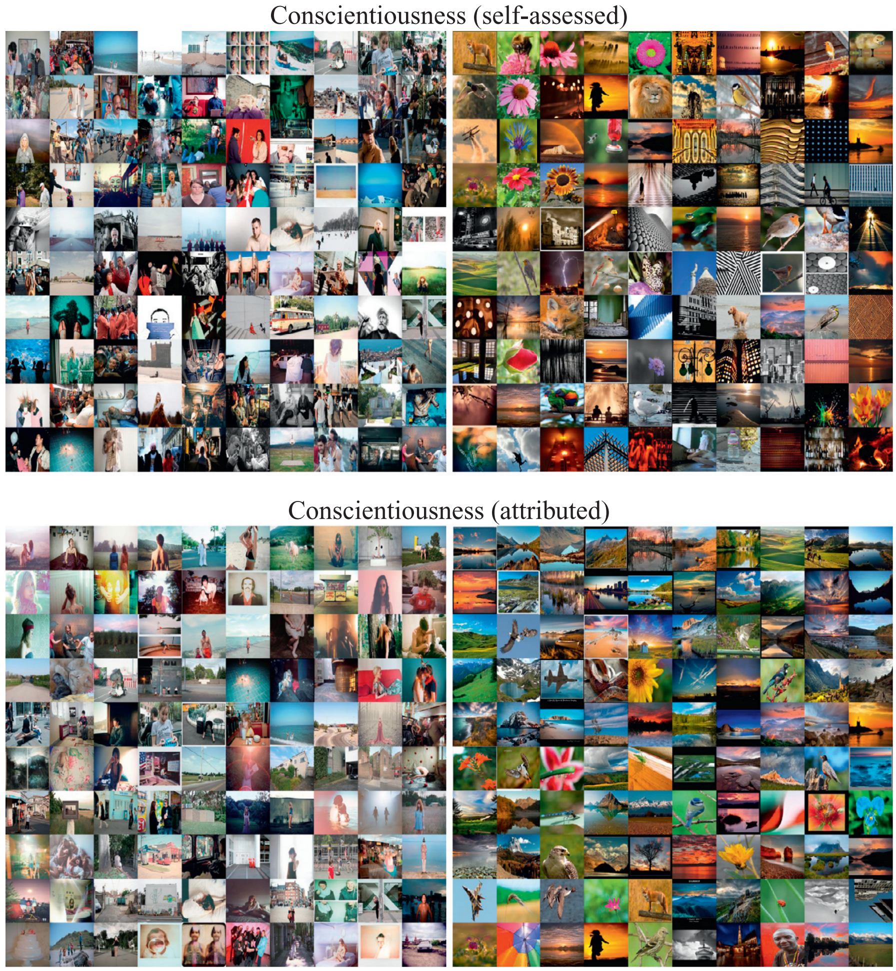


Fig. 7. Representative images in the low (left) and high (right) value classes for the Consciousness trait.

ated with the predictions, we select the top 9 correctly classified test images, average their feature maps to obtain a representative invariant map (a sort of archetype) and feed it through the fine-tuned net to obtain the associated probability estimate. We then selectively *kill* feature units in these maps to rate the effects on the estimate. By greedily removing the units starting from the one that lowers the probability the most if killed, we have a ranking of units that affect the prediction, that is, we find the features that most positively affect the prediction.

For example, if we consider low attributed Neuroticism, the average feature map of the top 9 correctly classified images is predicted as belonging to the class with probability 0.9825, higher even than the mean probability of the images themselves at 0.9479. Of the 512 feature units in this map, after killing only 41 of them, the probability drops below 0.5 (shown in Fig. 6): these 41 units are arguably meaningful for this class.

That list of units can then be used for the second task, reconstructing feature maps in the image space by using network deconvolution (Zhou et al., 2014a). Tables 5–8 show the top 5 most significant feature units for each class in terms of activation images, where visual reconstruction isolates only the relevant part in the input images.

For low attributed Neuroticism we can see that Unit 1 is specialized in detecting water in natural scenery (lakes, rivers etc.), Unit 2 and 3 are focused on yellow and red primarily, Unit 4 individuates curvy patterns while Unit 5 seems to focus on textures. In the high class, the first two units focus on flat uniform regions (not edgy), while Unit 3 on people (interestingly, ignoring the faces). Unit 4 on grayish/bluish patterns and Unit 5 on highly contrasted patterns.

In Table 5, at the bottom, we also show those 8 images that have been correctly classified with the highest probability and the 8 images that have been misclassified with the highest confidence. Correctly classified low Neuroticism images actually exhibit natural scenery with water and hot colors with curvy patterns. Correctly classified high Neuroticism images show many empty regions uniformly colored (grayish mostly) some of them with people inside, where the face is out of focus or severely shadowed. In many cases, there are examples of highly contrasted images. Incorrectly classified low Neuroticism images (they are pictures attributed as suggesting high Neuroticism) seem to be outliers, portraying colored natural scenery. Incorrectly classified high Neuroticism images, even in this case, share many aesthetic aspects with high Neuroticism, with the apparent difference that in most of the cases they portray children, which is an aspect that none of the units captured.

When it comes to Conscientiousness, shown in Fig. 7, we can observe for the high class the presence of orderly images, scene composed of landscapes, mountains where the line between the sky and floor is sharp or the presence of buildings as matter of appreciation for regular geometry, otherwise if there is an object in the image, it appears in the foreground in front of a blurred

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

Fig. 8. Representative images in the low (left) and high (right) value classes for the Extraversion trait.

background. In the *low* class we can see pictures where colors are faded, with pastel colors.

Concerning low attributed Conscientiousness, Unit 1 focuses on gray background regions, while Unit 2 on green areas. Unit 3 focuses on outdoor pavements, Unit 4 on monochromatic natural scenery and Unit 5 seemingly on faces. On attributed high conscientiousness, Unit 1 on the sky, Unit 2 on water scenery, as so as Unit 3, Unit 4 on squared regions, seemingly tv sets. Finally, Unit 5 on yellowish regions. In Table 6, correctly classified low Conscientiousness images show people (and faces) on mostly grayish background, while correctly classified high Conscientiousness images exhibit open scenarios with water and sky of primary importance, with hot colors. Incorrectly classified low Conscientiousness images and incorrectly classified high Conscientiousness images share strong aesthetic and content-based similarity with the opposite class, witnessing that the classification task is indeed difficult due to overlap between the two classes.

The Extraversion trait is the most impressive, shown in Fig. 8. For the high class we can see that all pictures are crowded of people, while the low class represents images where there are just flowers, plants, and indoor scenes, showing the tendency of introvert people to not interact as extrovert ones.

On low attributed Extraversion, reported in Table 7, Unit 1 focuses on very fragmented patterns that may indicate both flowers and animalsâÇÖ wings, while Unit 2 clearly on flowers. Unit 3 individuates curvy repeated patterns, Unit 4 desaturated scattered backgrounds and Unit 5 on repeated local curvy patterns. Attributed high Extraversion is very interesting, since all the units operate on images that exhibit low aesthetic quality, meaning no rule of third taken into account, no clear and unique subject captured by the camera, so in general the units seem to focus on aspects of images taken by normal people (not photographer, unless few exceptions). This may indicate that this trait is associated to people that take many pictures, without caring about the aesthetic quality. In particular, Unit 1 focused on text, Unit 2 on desaturated scenery with red insertions, Unit 3 and 4 on bluish and greenish details, respectively, and, finally, Unit 5 seems to capture accessories to be carried out when in an outdoor scenarios (cups, caps, newspapers). Correctly classified low Extraversion images show flowers and animals with mostly desaturated colors and curvy patterns while correctly classified high Extraversion images are amateur pictures of people in urban scenery. Incorrectly classified low Extraversion images (they belong to high Extraversion attributed people) share most of the characteristics (both concerning the aesthetic aspects and the content) of the opposite class. Incorrectly classified high Extraversion images apparently portray many people captured by improvised photographers, but it is very interesting to note that their expression is not happy in most of the cases (while this is frequent in high Extraversion attributed images), for example because they are protesting. Very probably this aspect has not been captured by any unit.

The Openness trait also shows really interesting emerging characteristics, shown in Fig. 9: high class show really complex and artistic pictures, ranging from black/white pictures to portraits and complex shapes and scene to abstract and surrealist pictures. the low class shows pictures related to animals, most of them are fe-

ARTICLE IN PRESS
JID: YCVIU [m5G;November 3, 2016;12:5]

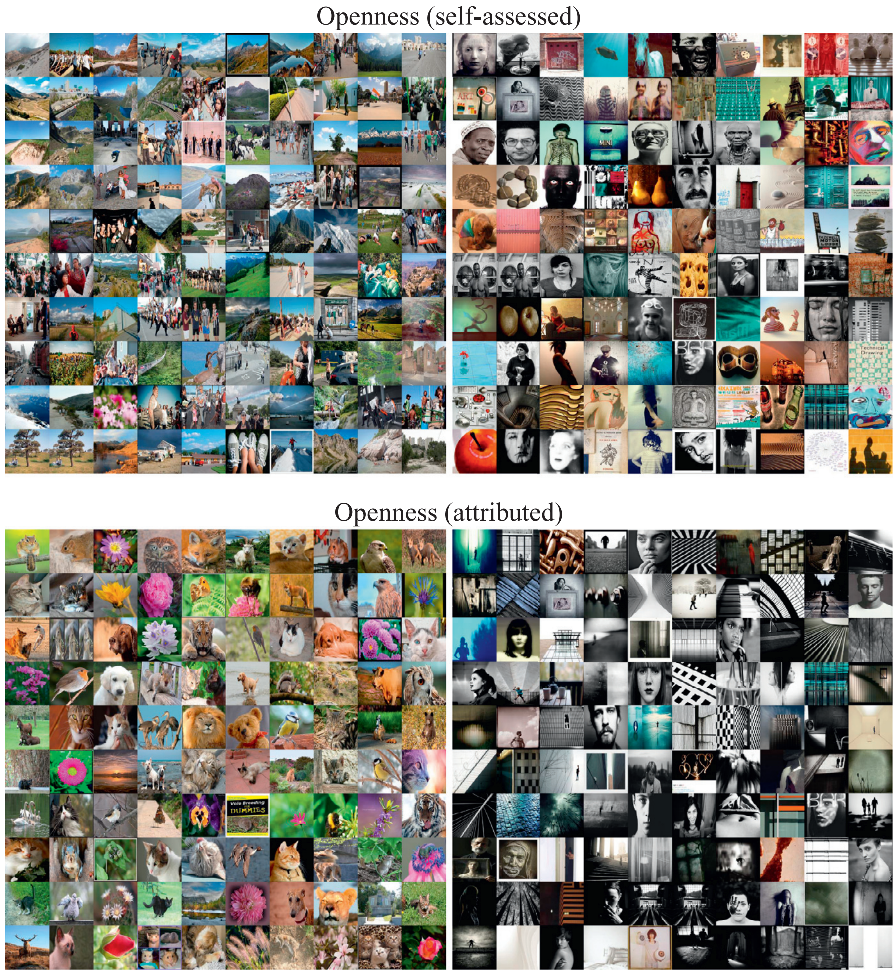


Fig. 9. Representative images in the low (left) and high (right) value classes for the Openness trait.

lines and flowers, highlighting the fact that this people tend to prefer simple, real life pictures.

Low attributed Openness, shown in Table 8, sees Unit 1 as clearly modeling cats, Unit 2 curvy patterns and flowers, Unit 3 natural scenery (unless few exception, few people inside), Unit 4 close ups of cats, and Unit 5 green animal faces. High attributed openness has all the units focusing in general on strongly geometrical patterns. Correctly classified low openness images show flowers and animals with mostly desaturated colors and curvy patterns while correctly classified high openness images are portraying strongly geometrical patterns. Incorrectly classified low openness images and incorrectly classified high openness images testify how difficult is our task, because they appear to belong to the opposite class.

Agreeableness high class show really colorful, high contrast, warm color pictures. On the other side, low class is represented by images mostly in black/white, and not friendly environments. Notice how there are quite a few images in common between attributed Neuroticism and Agreeableness, but in opposite classes. This is a confirmation of the strong negative correlation between the two attributed traits.

Attributed low Agreeableness is modeled by units that focus in general on desaturated images, as shown in Table 9, capturing cluttered backgrounds, without focusing on faces. High Agreeableness has Unit 1 focusing on water on natural scenarios, yellowish and reddish patterns (Unit 2 and 3, respectively) curvy colored local and repeated patterns (Unit 4) and umbrellas (Unit 5). Correctly classified low Agreeableness images show grayish and desaturated images with very few faces, while correctly classified high Agreeableness images are portraying natural scenery with repeated patterns (the petals). Incorrectly classified low Agreeableness images share many aspects with the low Agreeableness class, while incorrectly classified high Agreeableness images (they belong to the attributed low Agreeableness class), show actually natural scenarios as the high Agreeableness class, but they exhibit a slight tendency toward desaturated colors, and to have a lower aesthetic quality (no rule of thirds, unbalanced content).

In general, the very important add-on given by CNNs, is that the features encoded by the neural units are very different from all those one can find on the computational aesthetics literature. Very likely, a much larger dataset and/or more complex network architectures could succeed in encapsulating visual aspects we failed to grasp (the presence of children, facial expressions, rule of thirds, quoting the images that we analyzed in the previous paragraphs).

Lastly, we propose a deeper analysis of the semantic content learned for each trait. We take the high probability images of each class and perform unsupervised clustering using t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton, 2008). t-SNE is a technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets by giving each datapoint a location in a two or three-dimensional map. It is a way of converting a high-dimensional

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

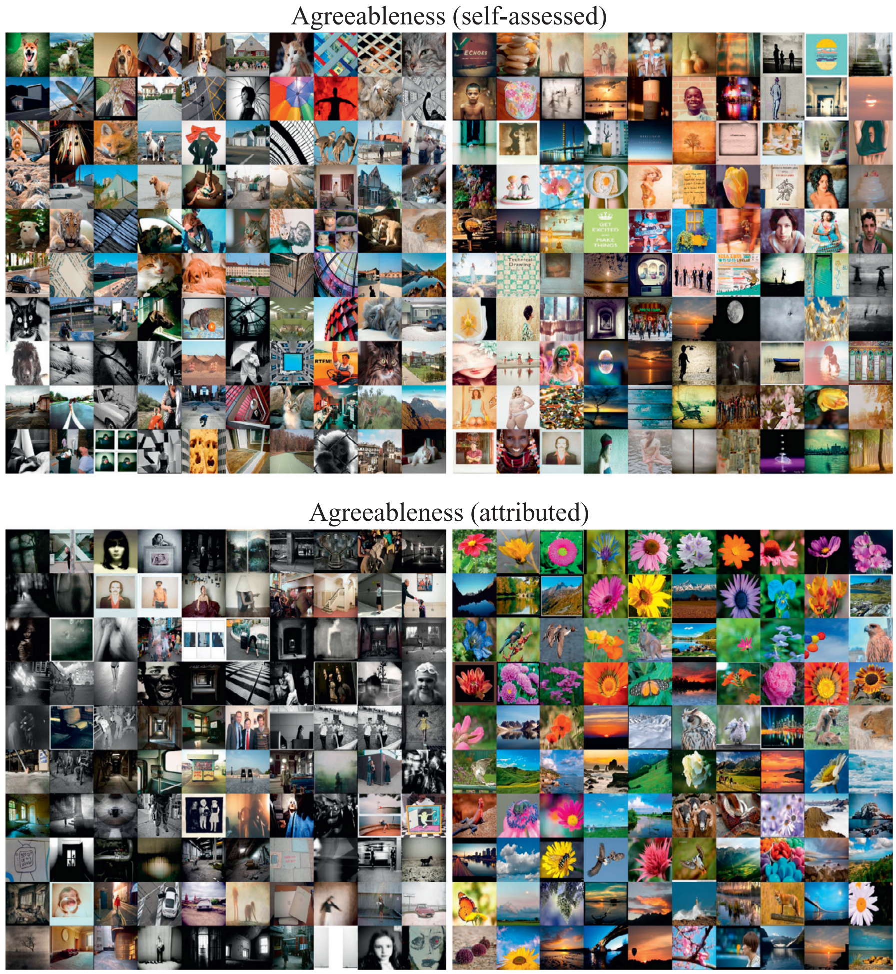


Fig. 10. Representative images in the low (left) and high (right) value classes for the Agreeableness trait.

data set into a matrix of pairwise similarities visualizing them and capable of capturing much of the local structure of the high-dimensional data very well, while also revealing global structure such as the presence of clusters at several scales. This applies well in our case where the features of the CNN are 4096 dimensional. The aim of dimensionality reduction is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map.

In our case, we use it to cluster and visualize how the most influential images of a given trait are grouped. This allows us to explore the semantics and attributes that describe the trait. To this purpose we retained the CNN features in the fully connected layer before the classification and gave them as input to t-SNE. The code preprocesses the data using PCA, reducing its dimensionality to a chosen dimensions (we cross-validated this parameter keeping values between 5 and 10). The perplexity basically sets how many near neighbors each point is trying to stay close to in the map. Therefore, small values for the perplexity will result in large numbers of small clusters, whereas large values will produce a smaller number of larger clusters. In our case we cross-validated this parameter as well achieving better and more interpretable results with a value of 5. The function returns a matrix that specifies the coordinates of the 2 low-dimensional data points. With this map we can build a plot where we can project each image on 2-dimensional space, allowing us to visualize the clusters.

In the specific for a high level of Neuroticism (see Fig. 11 top) we found three clusters about portrait of a person with blurred background, indoor scene and alone person in the dark. When it comes to a low level of Neuroticism (see Fig. 11 bottom) we found five clusters about birds, flowers, sunset/sunrise scene, landscape of mountains and animals in more general. In the case of high level of Openness (see Fig. 12 top) we found six clusters about portraits where the person is in a certain pose and nude, pastel colors paintings, pastel color shapes, artificial and digital painting, complex and twisted shapes similar to tissues, other artistic paintings. Instead for a low level of Openness (see Fig. 12 bottom) we found four clusters about flower/plants, felines, outside activity/sport and landscapes.

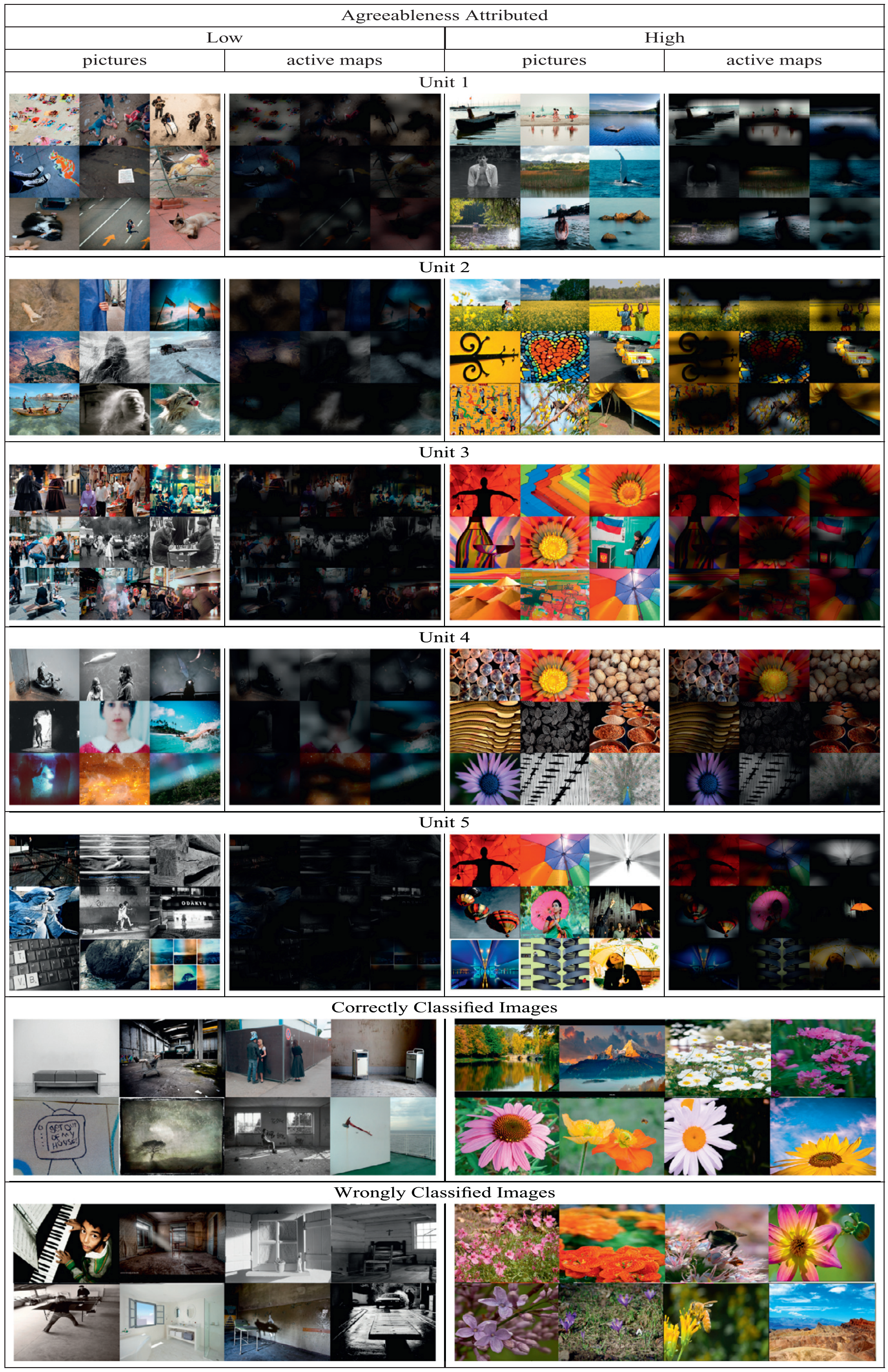
5. Demo

As a matter of proof that our proposed method effectively works, we developed a web interface where a subject can upload an image or a set of images, paste a web address linked to a picture or select an image form a list already stored in the server that he/she likes[1]. The proposed demo loads the models of the aesthetic preferences related to the attributed traits and classifies the pictures assigning them to the low or high class of each trait. If a set

[1] The demo is available at http://psychoflickr.di.univr.it:8000/demo/

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

Table 9
Display of the 5 most significant visual attributes for attributed Agreeableness, by montages of 9 images that strongly activate the corresponding feature units, together with activation images. The bottom rows show sets of correctly and wrongly classified images.



of images is provided, then it computes a majority vote to decide the level of the trait.

6. Conclusions

In this paper, we examine the problem of relating a set of image preferences to personality traits by using a deep learning framework. We cast this recently introduced application problem as a new level of image understanding that enhances the role of images through considerations on the social aspects of contemporary online activities. The role of social platforms like Flickr, Facebook, Instagram, etc., in building online social personas where most activities are shared to a wide audience creates a unique opportunity to study image-based activities (like authoring, uploading and

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

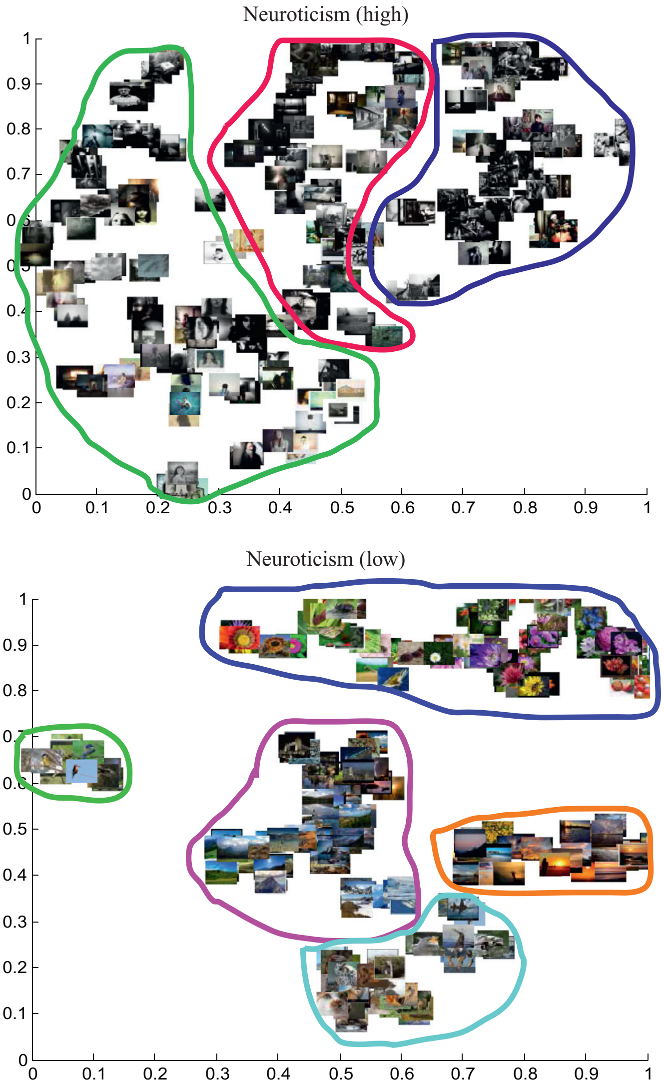


Fig. 11. Projection of the most probable images for Neuroticism trait, on a low-dimensional space.

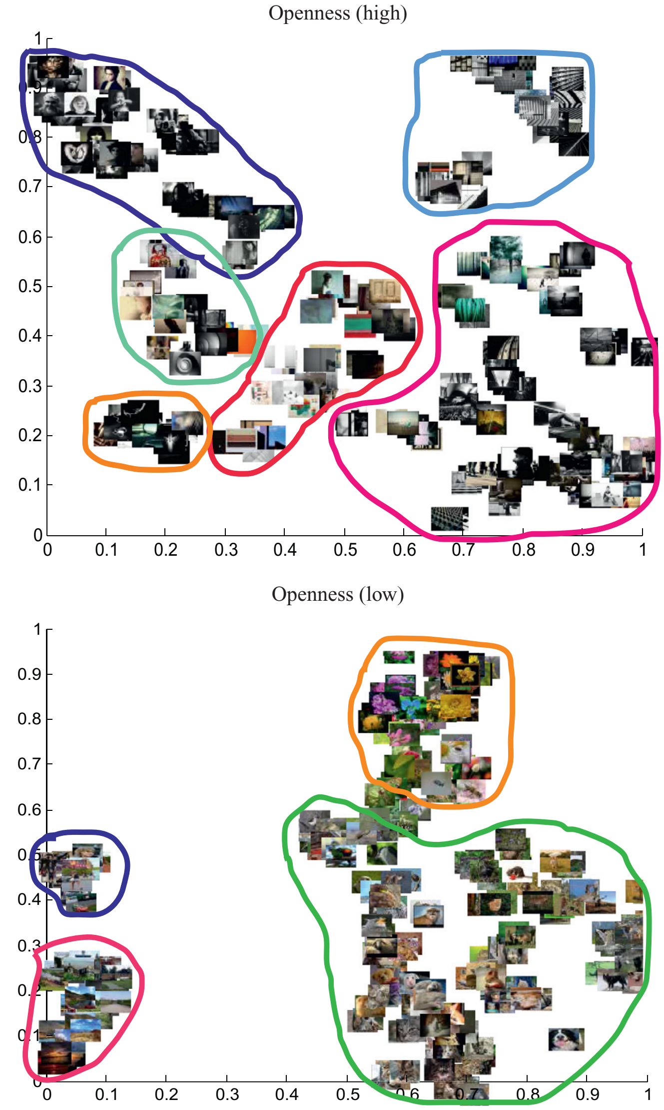


Fig. 12. Projection of the most probable images for Openness trait, on a low-dimensional space.

preferring images) as social messages, embedded with characteristics resembling verbal and non-verbal signals in face-to-face spoken communications.

Moreover, the presence of an author-audience paradigm imbues the messages with an extra layer of significance: there is the communication intended by the author, and there is the communication assumed by the audience. Thus, our problem becomes inferring both self-assessed and attributed personality traits, and our results reinforce previous work in considering the latter easier to approach than the former.

Overall, we demonstrate the viability of using a recent, powerful methodology like convolutional neural networks, in tackling these new types of image understanding applications. The experimental results are promising, outperform previous results and point towards a mixture of stylistic aspects (typical of computational aesthetics) and content-based aspects (typical of object detectors) as crucial for building reliable predictors. We believe there are many impactful rewards for this type of research: an immediate application might be providing social networks with tools to soft-profile users, suggesting compatible users to connect with, or indicating the most suitable groups to join. It could also be used as a marketing evaluation tool to help predict the impact of a set of images on an hypothetical audience of customers. Our online demo is a step along this direction, evidence that this new kind of image understanding is not only a mere academic research endeavor, but a potential groundbreaking market application.

Acknowledgment

Dong Seon Cheng was supported by the Hankuk University of Foreign Studies Research Fund of 2015.

References

Bauckhage, C., Kersting, K., 2013. Can computers learn from the aesthetic wisdom of the crowd? KI-Künstliche Intelligenz 27 (1), 25–35.
Birkhoff, G.D., 1933. Aesthetic measure.
Brunswik, E., 1956. Perception and the Representative Design of Psychological Experiments. University of California Press.
Campbell, A., Ciesielksi, V., Qin, A., 2015. Feature discovery by deep learning for aesthetic analysis of evolved abstract images. In: Evolutionary and Biologically Inspired Music, Sound, Art and Design. Springer, pp. 27–38.
Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J., 1999. Blobworld: A system for region-based image indexing and retrieval. In: Visual Information and Information Systems. Springer, pp. 509–517.
Ciesielski, V., Barile, P., Trist, K., 2013. Finding Image Features Associated with High Aesthetic Value by Machine Learning. Springer.

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013

Costa, P.T., MacCrae, R.R., Psychological Assessment Resources, I., 1992. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual. PAR.
Cristani, M., Vinciarelli, A., Segalin, C., Perina, A., 2013. Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In: Proceedings of the ACM international conference on Multimedia. ACM, pp. 213–222.
Datta, R., 2009. Semantics and Aesthetics Inference for Image Search: Statistical Learning Approaches. Ph.D. thesis. University Park, PA, USA.
Datta, R., Joshi, D., Li, J., Wang, J., 2006. Studying aesthetics in photographic images using a computational approach. In: ECCV, pp. 288–301.
Dhar, S., Ordonez, V., Berg, T.L., 2011. High level describable attributes for predicting aesthetics and interestingness. In: CVPR, pp. 1657–1664.
Furnham, A., Avison, M., 1997. Personality and preference for surreal paintings. Pers. Individual Differ. 23 (6), 923–935.
Furnham, A., Walker, J., 2001. The influence of personality traits, previous experience of art, and demographic variables on artistic preference. Personality and Individual Differences.
Galanter, P., 2012. Computational aesthetic evaluation: past and future. In: Computers and Creativity. Springer, pp. 255–293.
Golbeck, J., Robles, C., Edmondson, M., Turner, K., 2011. Predicting personality from twitter. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, pp. 149–156.
Gong, E.,. Deep aesthetic learning.
Gosling, S.D., Gaddis, S., Vazire, S., et al., 2007. Personality impressions based on facebook profiles.. ICWSM 7, 1–4.
Hoenig, F., 2005. Defining computational aesthetics. In: Proceedings of the Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 13–18.
Hughes, D.J., Rowe, M., Batey, M., Lee, A., 2012. A tale of two sites: twitter vs. facebook and the personality predictors of social media usage. Comput. Human Behav. 28 (2), 561–569.
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093.
Jin, X., Wang, C., Luo, J., Yu, X., Han, J., 2011. Likeminer: a system for mining the power of 'like' in social media networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 753–756.
Jojic, N., Perina, A., 2011. Multidimensional counting grids: Inferring word order from disordered bags of words. In: Proceedings of the Twenty-Seventh Conference Annual Conference on UAI. AUAI Press, Corvallis, Oregon, pp. 547–556.
Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.-T., Wang, J.Z., Li, J., Luo, J., 2011. Aesthetics and emotions in images. Signal Process. Mag. IEEE 28 (5), 94–115.
Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H., 2013. Recognizing image style. arXiv preprint. arXiv:1311.3715.
Ke, Y., Tang, X., Jing, F., 2006. The design of high-level features for photo quality assessment. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 1. IEEE, pp. 419–426.
Kosinski, M., Stillwell, D., Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. Proc. National Acad. Sciences 110 (15), 5802–5805.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.
Li, L.-J., Su, H., Fei-Fei, L., Xing, E.P., 2010. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In: Advances in Neural Information Processing Systems, pp. 1378–1386.
Liu, Y., Zhang, D., Lu, G., Ma, W.-Y., 2007. A survey of content-based image retrieval with high-level semantics. Pattern Recognit. 40 (1), 262–282.
Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.Z., 2014. Rapid: Rating pictorial aesthetics using deep learning. In: Proceedings of the ACM International Conference on Multimedia. ACM, pp. 457–466.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (2579–2605), 85.
Machado, P., Cardoso, A., 2010. Generation and evaluation of artworks. In: Proceedings of the 1st European Workshop on Cognitive Modeling, CM, Vol. 96, pp. 39–96.
Murray, N., Marchesotti, L., Perronnin, F., 2012. Ava: A large-scale database for aesthetic visual analysis. In: CVPR, pp. 2408–2415.
Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J., 2011. Our twitter profiles, our selves: Predicting personality with twitter. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, pp. 180–185.
Quercia, D., Lambiotte, R., Stillwell, D., Kosinski, M., Crowcroft, J., 2012. The personality of popular facebook users. In: Proceedings of the ACM 2012 conference on computer supported cooperative work. ACM, pp. 955–964.
Rammstedt, B., John, O.P., 2007. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. J. Res. Pers. 41 (1), 203–212.
Rawlings, D., Twomey, F., Burns, E., Morris, S., 1998. Personality, creativity and aesthetic preference: comparing psychoticism, sensation seeking, schizotypy and openness to experience. Empirical Stud. Arts 16 (2), 153–178.
Segalin, C., Cristani, M., Perina, A., Vinciarelli, A., 2016. The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits. IEEE Trans. Affect. Comput. PP (99) 1–1.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint. arXiv:1409.1556.
Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years. Pattern Anal. Mach. Intell. IEEE Trans. 22 (12), 1349–1380.
Tibshirani, R., 1994. Regression shrinkage and selection via the lasso. JRSS 58, 267–288.
Vailaya, A., Figueiredo, M.A., Jain, A.K., Zhang, H.-J., 2001. Image classification for content-based indexing. Image Process. IEEE Trans. 10 (1), 117–130.
Vedaldi, A., Lenc, K., 2015. Matconvnet – convolutional neural networks for MATLAB.
Vinciarelli, A., Mohammadi, G., 2014. A survey of personality computing. IEEE Trans. Affect. Comput. 5 (3), 273–291.
Vinciarelli, A., Pentland, A.S., 2015. New social signals in a new interaction world: the next frontier for social signal processing. Syst. Man Cybern. Mag. IEEE 1 (2), 10–17.
Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2014. Object detectors emerge in deep scene cnns. arXiv preprint. arXiv:1412.6856.
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495.

Please cite this article as: C. Segalin et al., Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.013