



Citation for published version:

Wei, X & Stillwell, D 2017, How smart does your profile image look? Estimating intelligence from social network profile images. in *Proceedings of the 10th ACM international conference on Web search and data mining*. Association for Computing Machinery, New York, U. S. A., pp. 33-40. <https://doi.org/10.1145/3018661.3018663>

DOI:

[10.1145/3018661.3018663](https://doi.org/10.1145/3018661.3018663)

Publication date:

2017

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

How smart does your profile image look? Estimating intelligence from social network profile images

Xingjie Wei^{1,2}
x.wei@bath.ac.uk

David Stillwell¹
ds617@cam.ac.uk

Psychometrics Centre, University of Cambridge, Cambridge, CB2 1AG, U.K.¹
School of Management, University of Bath, Bath, BA2 7AY, U.K.²

ABSTRACT

Profile images on social networks are users' opportunity to present themselves and to affect how others judge them. We examine what Facebook images say about users' perceived and measured intelligence. 1,122 Facebook users completed a matrices intelligence test and shared their current Facebook profile image. Strangers also rated the images for perceived intelligence. We use automatically extracted image features to predict both measured and perceived intelligence. Intelligence estimation from images is a difficult task even for humans, but experimental results show that human accuracy can be equalled using computing methods. We report the image features that predict both measured and perceived intelligence, and highlight misleading features such as "smiling" and "wearing glasses" that are correlated with perceived but not measured intelligence. Our results give insights into inaccurate stereotyping from profile images and also have implications for privacy, especially since in most social networks profile images are public by default.

Keywords

Intelligence quotient; Measured intelligence; Perceived intelligence; Intelligence estimation; Computational aesthetics; IQ

1. INTRODUCTION

Profile images are fundamental to online social networks. They are typically displayed each time a user posts a message or shares a piece of content, and are normally public by default. As such, they are an important avenue for users to share their self-representation, and they can have a big effect on how friends and strangers judge the user. Recent experimental research [2] shows that Facebook profile images affect employers' hiring decisions: Candidates with the most attractive profile images obtained 39% more job interview invitations than those with the least attractive images.

General profile images include behavioural cues such as

camera views, poses, clothes, locations and the presence of friends and partners; or preference cues such as pets, cartoon characters, sports equipment and special signs, which are linked to people's lifestyles and how they represent themselves in the social network. Such cues are the outcome of intentional choices, driven by psychological differences.

Intelligence is an individual difference that is related to important life outcomes such as income [27] and relationships [9]. The first impressions of intelligence can have significant consequences in social scenarios such as employment. Different from other *neutral* psychological traits such as personality, where there is no ideal personality, high intelligence is a trait that people want to project to others by self-representation.

In this paper, we investigate whether a user's intelligence can be inferred from profile images. This will help us to better understand human behaviour, help Web users to better manage their self-representation, and enable an understanding of the privacy risk of having a public profile image. There are very few works studying the relationship between images of people and their intelligence [10, 23, ?], and those that exist focus on facial photographs using small sample sizes. Our work is distinguished from those by taking into account the general profile images which contain more psychological and behavioural cues, such as the image background, and using a much larger data set.

We define a user's intelligence score measured by an IQ test as *measured intelligence* (MI) and the intelligence score rated by human observers' perceptions based on the self-representation of users as *perceived intelligence* (PI). We propose a framework of intelligence estimation from profile images by both humans and computers as shown in Figure 1. We use profile images from 1,122 Facebook users who have taken an IQ test, and whose profile images have also been rated by strangers. We extract visual features from profile images and examine the relationships between those features and both MI and PI. Our contributions can be summarised as follows:

- We designed experiments to investigate how humans make intelligence judgments (i.e., PI) for others from profile images. We found different people have relatively consistent judgements and such judgements are associated with MI for both men and women.
- We proposed a system of automatic intelligence estimation from profile images and evaluated the estimation accuracy of both computers and humans. We found it is possible to equal humans' accuracy using

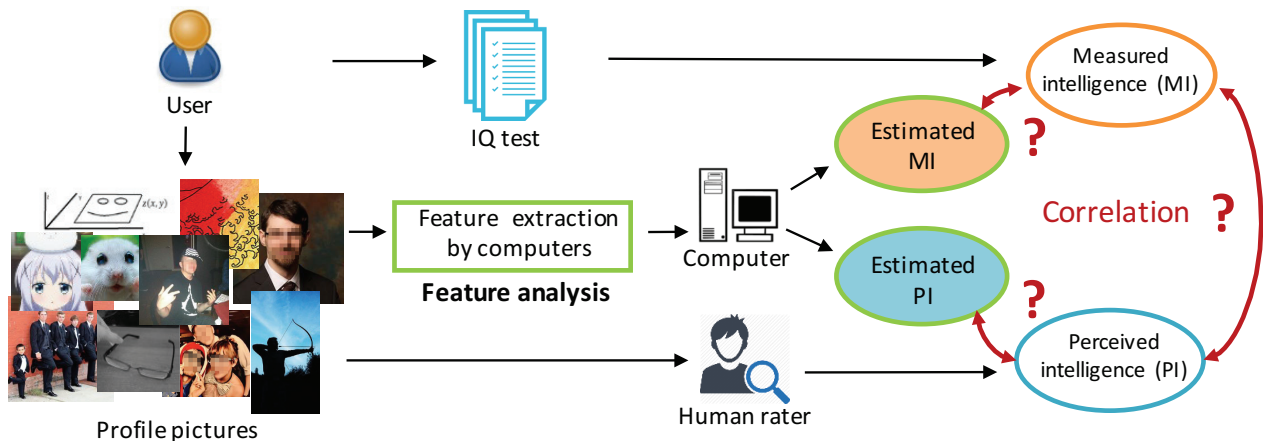


Figure 1: Illustration of image based intelligence estimation.

algorithms while also having the potential to reduce biased judgements.

- We conducted comprehensive analysis on the relationships between visual features and MI or PI. We answered the research questions about what visual elements an intelligent person will use in profile images, and what visual elements in a profile image make a person perceived to be intelligent. Those insights can guide people to manage their self-representations.

The rest of this paper is organised as follows: Section 2 reviews related works in both psychological and computing literatures. Section 3 introduces the proposed framework including how the MI score is generated and how the MI and PI are estimated by humans and computers. Section 4 presents the experimental results of feature analysis and intelligence estimation. Section 5 draws the conclusions and finally Section 6 summarises the future research directions.

2. RELATED WORK

Recent research [12] has examined the extent to which seemingly innocuous social media data such as Facebook status updates [20] and Facebook ‘Likes’ [26] can be used to infer highly intimate behavioural and psychological traits. Compared to using other types of social media data such as text [20] for behaviour analysis, the processing of image data is less studied due to the challenge of extracting predictive features (see the survey in [25]). Redi *et al.* [18] predicted the ambiance of commercial places based on the profile pictures of its visitors in Foursquare. Guntuku *et al.* [8] predicted users’ preference for an image based on its visual content and associated tags. Steele *et al.* [22] explored the human raters’ agreement of personality impression from profile images. Liu *et al.* [13] analysed how Twitter profile images vary with the personality of the users. Cristani *et al.* [6] and Segalin *et al.* [21] predicted both measured and perceived personality scores using Flickr images tagged as favourite by users.

There are very few works studying the relationship between images of people and their intelligence. Kleisner *et al.* [10] used geometric morphometrics to determine which facial traits are associated with MI and PI based on 80 facial photographs. Results shown that certain facial traits were

shown to be correlated with PI while there was no correlation between morphological traits and MI; PI was associated with MI in men but not women. Talamas *et al.* [23] investigated PI, attractiveness and academic performance using 100 face photos and found no relationship between PI and academic performance, but a strong positive correlation between attractiveness and PI. They also scrutinized PI and attractiveness in the work in [?] based on face photos of 100 adults and 90 school-aged children and found that PI was partially mediated by attractiveness, but independent effects of centre facial cues such as eyelid-openness and subtle smiling can enhance PI scores independent of attractiveness.

Compared with the above works, we employ social network profile images which include more behavioural cues linking to people’s lifestyles, interests, values and thoughts, not just limited to facial cues such as attractiveness or symmetry. We investigate and predict both MI and PI, and also consider demographic information such as gender and age in the analysis. To the best of our knowledge, our work is the first one to automatically estimate intelligence from profile images in social network.

3. PROPOSED FRAMEWORK

Our framework contains three main elements: 1) users take an IQ test to get their MI scores, 2) users’ profile images are shown to other human raters to get PI scores, and 3) visual features are algorithmically extracted from images to estimate MI and PI separately.

3.1 Data collection

Our data is based on the *myPersonality*¹ database. *myPersonality* was a Facebook application where users can take psychometric tests by logging in using their Facebook account. Now the database contains more than 6 million psychometric test results together with more than 4 million individual Facebook profiles. About 7,000 users took a 20-item matrices IQ test (validated against Raven’s Matrices [17]) designed by professional psychometricians; the MI scores of those users are calculated from the IQ test. Their Facebook profiles were collected with consent, which contain demographic information such as gender and age.

¹<http://mypersonality.org/>

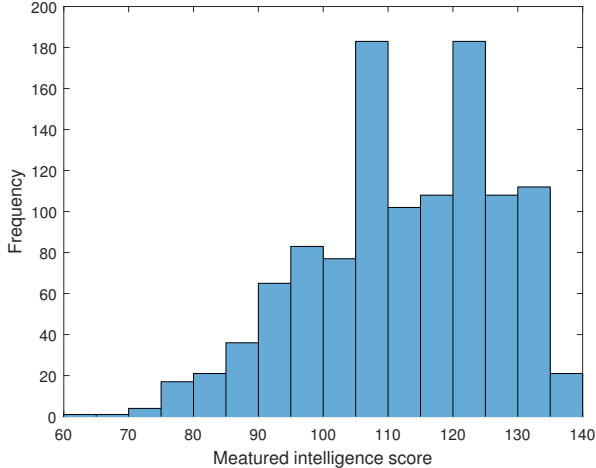


Figure 2: Histogram of MI scores. The collected samples showed a broad range of distribution of MI scores.

We selected those users who shared their profile images (excluding the Facebook default profile image) and there are 1,122 users (51% men, age mean \pm std=25.9 \pm 9.2, range: 14~69; MI score mean \pm std=112.4 \pm 14.5, range: 64.9~138.6, distribution is shown in Figure 2). Each user has one publicly available profile image in JPEG format, normally of size 200 \times 200 pixels. The thumbnails of some example images are shown in Figure 1. There are about 16% non-person images (e.g., cartoons, drawings, animals, signs, etc.), 60% with only one person, 21% with two or three persons and 3% group images (more than four persons). Images with persons may also contain other content such as animals, plants, sports, indoor and outdoor scenes, events (e.g., wedding or graduation), etc.

We also recruited 739 independent raters (49% men, age mean \pm std=24.2 \pm 6.2, range: 15~72 covering a broad range of raters with different ages) through our online psychometrics platform Concerto². Each rater was randomly shown 50 or 100 images, one image at each time. Raters received the following instructions:

At each time, a profile image will be presented to you. The image content may contain (one or more than one) people, animals or other objects. You will be asked to make a judgement of how intelligent this user is who uses such an image as their social network profile image. You need to rate the intelligence score for a given image using a 7-scale point scale wherein 7 stands for the highest ranking (most intelligent) and 1 the lowest (least intelligent). There is no time limit for rating each image.

Each image was finally rated by at least 24 different raters. Each rater’s scores were z-score-normalised in order to account for positivity or negativity bias.

3.2 Perceived intelligence estimation by humans

Figure 3 shows an example of the rated scores for the first 20 images. Each box indicates the distribution of rated

²<http://concertoplatform.com/>

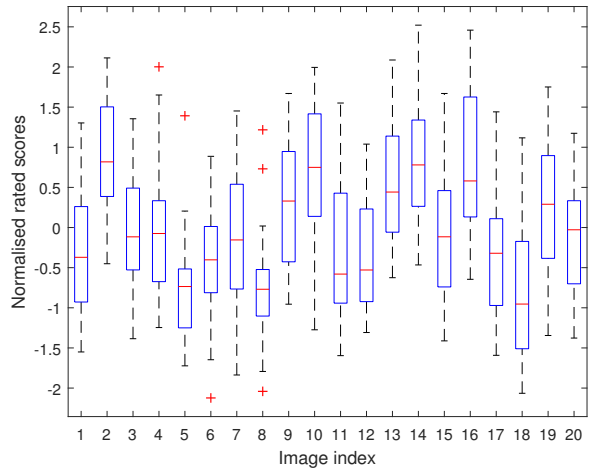


Figure 3: Rated scores for the first 20 images.

scores for each image. The box length is the interquartile range of the scores and the red line indicates the median value. Red crosses indicate some extreme scores where raters strongly disagree with the average. Whiskers indicate the maximal and minimal scores (excluding extreme scores). For each image, there are few extreme scores in the plot and most rated scores tend to cluster in a relatively narrow range. This example shows raters’ PI scores are relatively consistent within images but there are differences between images.

In order to quantitatively evaluate whether different raters agree with each other, we compute the inter-rater reliability (IRR) [19] for rated scores. The rated scores (observed scores) X is composed of a true score T and a measurement error (noise) E between different raters [16]. The IRR analysis aims to determine how much of the variance of the rated scores is due to variance in the true scores after removing the variance of measurement error:

$$Reliability = \frac{var(T)}{var(T) + var(E)} \quad (1)$$

To rate all 1,122 images by each rater is time-consuming and raters may lose attention due to long time work. As mentioned in Section 3.1, in our experimental settings, each rater randomly rated 50 or 100 images and each image was finally rated by at least 24 different raters. We use the interclass correction coefficient (ICC) [16] to measure the IRR because ICC is suitable when a subset of samples is rated by multiple raters. It models the rated score X_{ij} of image i by rater j as:

$$X_{ij} = \mu + r_i + e_{ij}, \quad (2)$$

where μ is the unobserved overall mean which is constant; r_i is an unobserved random effect shared by all rated scores for image i and e_{ij} is an unobserved measurement error. r_i and e_{ij} are assumed to have expected value zero and to be uncorrelated with each other. The variance of r_j is denoted σ_r^2 and the variance of e_{ij} is denoted σ_e^2 . The degree of absolute agreement based on average of k rated scores ($k = 24$ in our data) is calculated using the ICC one-way model as

(more details see [16]):

$$ICC = \frac{\sigma_r^2}{\sigma_r^2 + \frac{\sigma_e^2}{k}} \quad (3)$$

Higher ICC values indicate greater IRR (guideline[5]: less than 0.4: *poor*, between 0.4 and 0.59: *fair*, between 0.6 and 0.74: *good*, between 0.75 and 1: *excellent*) - in this work it was 0.86. This means that most raters agree with one another in their perception of each image’s intelligence. Therefore, the PI score for each image (user) is calculated as the median value of its rated scores. We use median value instead of mean value because the former is more robust to the some extreme rated scores shown in Figure 3.

3.3 Intelligence estimation by computers

We model the intelligence estimation as a regression problem by mapping image features into MI or PI scores. In this section we first introduce the image features extracted and then present the feature selection schemes used for MI regression and PI regression. After feature selection, the regression is implemented using Support Vector Regression (SVR).

3.3.1 Feature extraction

We use the image aesthetics and perception descriptors [15, 4] inspired by psychology and art theories to extract useful features from profile images. Those features are understandable and effective in human behaviour analysis tasks as shown in previous works [6, 8, 21]. Those features can be classified into four main categories: colour, composition, texture, and body & face related features. The colour, composition and texture features capture low-level information while the body & face related features capture the high-level, i.e., content or semantic information. Body is detected using the Deformable Part Model [11] and Face related features are extracted by the *Face++* detector [28].

Additionally, as well as extracting the above features from a whole image, we also partition each image into 4×4 blocks then extract colour and texture features such as LBP [1] and dense SIFT [14], similar as the work in [7]. LBP is widely used to encode facial texture and dense SIFT is popular for representing scene images. Extracting those features from each block is able to capture the local structure of an image. Section 4.1 will show that there are a large number of such local features significantly correlated with both MI and PI, which can be very useful for intelligence estimation. Details of each feature are shown in Table 1.

3.3.2 Feature selection

After feature extraction, all features of an image are concatenated into an M -dimensional vector. We employ Principle Component Analysis (PCA) to reduce feature dimension by selecting N dimensions which have the highest amount of variance where N is the number of training samples ($N \ll M$).

Next, we adopt filter based feature selection on the uncorrelated PCA features to select the most informative features. In psychometrics tests, MI scores can be classified into several categories according to the values. For example:

- 1-very superior: ≥ 130
- 2-superior: 120~129
- 3-high average: 110~119

- 4-average: 90~109
- 5-low average: 80~89
- 6-borderline: 70~79
- 7-low: ≤ 69

Inspired by this, we choose the most discriminant features while constraining the same feature to be selected across different categories of MI scores. This is similar to the idea of multi-task learning [29] which learns models for multiple tasks jointly using a shared representation.

We perform univariate statistical tests on features and the target variable in the training set and select the most statistically significant K features (according to p-value/F-score by F-test). For MI, we select K features based on MI score and another K features based on MI label and then use the intersection of those two feature sets as the final selected features. For PI, the scores are already calculated based on a 7-point scale, so the target variable is the PI score directly.

4. EXPERIMENTS AND RESULTS

Parameters for image feature extraction are set as the optimal values reported in their original papers. We present univariate correlations between features and intelligence scores and then report the overall intelligence estimation results. Experiments are performed using the Leave-one-out cross-validation: Each time, one image is selected as testing sample and the remaining images are training samples on which regression models are trained. The selected dimension K in section 3.3.2 is set as $N/2$ where N is the number of training samples. We use cross-validation in training set to find the optimal parameters for SVR to avoid overfitting.

4.1 Feature analysis

We calculate the Spearman correlation coefficient ρ between each individual feature and MI or PI scores. In this section we describe: 1) what visual elements an intelligent person will use in profile images, and 2) what visual elements in a profile image make a person perceived to be intelligent.

As shown in Table 2, the ratios of significantly (p value < 0.05) correlated features with PI are much larger than that with MI. This might be because PI scores are based on only raters’ perceptions of a user’s profile image, but MI scores may involve other factors which cannot be directly inferred from images such as education level. Most of the colour, composition, body and face, and texture features are significantly correlated with MI and PI. Although the ratios of local features are relatively low, considering the high dimension of those features there are still a large number of features which are useful for regression.

Correlations between individual features and MI or PI are shown in Figure 4. For local features, since the dimensions are very high, only features with significant correlations are selected in the calculation and the average positive and negative correlations are shown. For face and body features, only images with faces are included.

4.1.1 Colour

The percentage of green pixels and colour sensitivity are positively correlated with MI while H circular variance (colour diversity in hue), percentages of pink, purple and red pixels, and dark channel (the higher the value, the less clear the image) are negatively correlated with MI. This suggests people

Table 1: Description of image features

	Category	Name	Len.	Description
Low-level	Colour	HSV statistics	5	Circular variance of H channel, average of S, V (use of light), standard deviation of S, V
		Emotion-based	3	<i>Valence</i> , <i>Arousal</i> and <i>Dominance</i> in V and S channels
		Colourfulness	1	Colour diversity
		Colour name	11	The percentage of black, blue, brown, grey, orange, pink, purple, red, white and yellow pixels
		Dark channel	1	The minimum filter output on RGB channel, reflects image clarity, saturation and hue
	Colour sensitivity	1	The peak of a weighted colour histogram representing the sensitivity with respect to human eye	
	Composition	Edge pixels	1	The percentage of edge pixels to present the structure of an image
		Regions	2	Number of regions, average size of regions
		Symmetry	2	Horizontal symmetry and vertical symmetry
	Texture	Entropy	1	Gray distribution entropy
		Sharpness	4	The average, variance, minimal and maximal value of sharpness
		Wavelet	12	Wavelet textures (spatial smoothness/graininess) in 3 levels on each HSV channel
		Tamura	3	Coarseness, contrast and directionality of texture
		GLCM	12	Contrast, correlation, energy, homogeneousness for each HSV channel
		GIST	24	Low dimensional representation of a scene, extracting from a whole image
	Local	Colour histogram	512	Histogram of colour from local blocks
LBP		944	Local Binary Pattern ($LBP_{i,2}^{u,2}$) from local blocks	
GIST		512	GIST features extracted from local blocks	
SIFT		2048	Dense SIFT features from local blocks	
High-level	Body & face	Body	2	The presence of body* and the proportion of the main body
		Skin	1	The percentage of skin pixels
		Face	4	The number of faces*, the proportion of main face, the horizontal and vertical locations of main face
		Glasses	2	The presence of normal glasses* or sunglasses*
		Pose	3	The pitch angle, roll angle and yaw angle of head

* with manual check to make sure the automatic detection results are correct

with high MI scores like to represent themselves in a clear and simple profile image. Percentages of grey and white pixels are positively correlated with PI, which suggests those colours in a profile image makes the user look intelligent. H circular variance, average of S channel (indicates chromatic purity, the lower the value, the purer the colour), percentages of brown, green, pink, purple pixels and dark channel are negatively correlated with PI.

4.1.2 Composition

Percentage of edge pixels is negatively correlated with PI. Most other composition features are correlated with neither MI nor PI. This suggests that in this dataset people do not prefer particular composition elements in their profile images. Moreover, those elements also do not affect how others judge the user's intelligence.

4.1.3 Body & face

The proportions of skin and face area are negatively correlated with MI. Similarly, the proportions of body, skin and face are negatively correlated with PI. Users who are very close to the camera are perceived as less intelligent. The smile degree and the presence of glasses are positively correlated with PI while no such correlations are observed in MI, indicating that people whose pictures show them smiling

and wearing glasses are perceived as being intelligent but this is an inaccurate stereotype.

4.1.4 Texture

The average, minimal and maximal values of sharpness are positively correlated with both MI and PI. The variance of sharpness is negatively correlated with PI. This confirms that the sharper (more clear in texture) a profile image is, the more intelligent the user is perceived to be. Most wavelet features, which represent graininess, are negatively correlated with MI. Most GIST features, which represent the dominant spatial structure of a scene, are positively correlated with both MI and PI.

4.1.5 Local

The average positive and negative correlations are shown separately. The texture features LBP achieve the highest correlations for both MI and PI.

4.2 Intelligence estimation

4.2.1 By humans

The previous research [23] which is based on 100 facial photographs showed that PI is associated MI in men (photo) but not women (photo). Here we also partition users into two groups by gender and analyse whether PI can predict

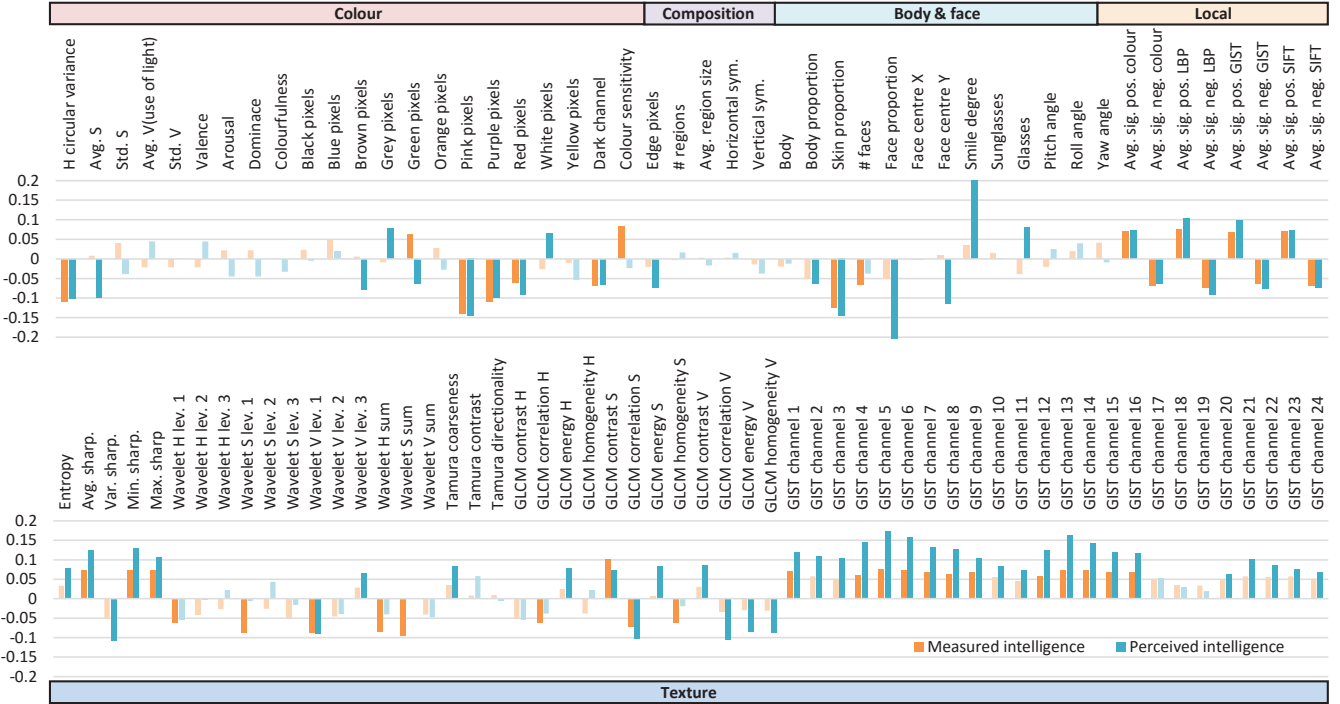


Figure 4: Correlations between image features and MI (orange bars) or PI (blue bars). Darker bars indicate correlations which are statistically significant ($p < 0.05$).

Table 2: Ratios of significantly correlated feature ($p < 0.05$)

Feature	Len.	$p < 0.05$ ratio	
		MI	PI
Colour, composition, body&face, texture	95	34.3%	56.3%
Local colour histogram	512	8.0%	11.9%
Local LBP	944	15.4%	44.6%
Local GIST	512	9.0%	43.2%
Local SIFT	2048	7.4%	11.7%

MI, i.e., to see is there any difference when human raters make intelligence judgement for male users and female users.

For each user group (i.e., male user or female user), we run the linear regression to test the association between MI (dependent variable) and PI (independent variable) while taking the age of users as controlled variable. We found that PI significantly affects MI in both male and female users (both $p < 0.001$). For better comparison, we report the correlation between PI and MI for male and female users separately as shown in Table 3. PI are significantly correlated with MI for both male and female users. This is because the work in [23] uses images containing only cropped faces and raters are more likely affected by attractiveness of the face. Our profile images contains more behavioural cues, as mentioned in Section 2, which provides more information for raters to make their intelligence judgements. But we also confirm the observation that the correlation for female users are lower than that for male users in all rater groups (i.e., male, female and together).

4.2.2 By computers

We report both the MI and PI estimation results in terms of correlation and root-mean-square error (RMSE) as shown in Table 4. Since MI and PI scores are in different scales,

Table 3: Spearman correlation ρ between PI and MI

	Male users	Female users	Together
Male raters	0.23	0.21	0.24
Female raters	0.23	0.18	0.22
Together	0.25	0.20	0.24

All correlations are significant at $p < 0.001$ level

we also report the Normalized RMSE to allow comparison (i.e., $RMSE / (y_{max} - y_{min})$ where y is the actual MI or PI scores). Similar to previous work [21], we compare the performance of human and computer with two methods which generate predicted scores from: 1) Random: random scores from a Gaussian distribution with the same mean and variance of training scores (we run this ten times and reported the average results), and 2) Mean: the average of training scores, which can minimise the RMSE.

Table 4: Estimation results

	Spearman ρ	RMSE	NRMSE
MI			
Human	0.24***	–	–
Computer	0.27***	14.50	0.20
Random	$< 0^*$	15.13	0.21
Mean	–	14.49	0.20
PI			
Computer	0.36***	0.54	0.15
Random	$< 0^{**}$	0.58	0.17
Mean	–	0.56	0.16

*** : $p < 0.001$, ** : $p < 0.01$, * : $p < 0.05$

For MI, the computer’s estimation ($\rho = 0.27$) is slightly better than that of humans ($\rho = 0.24$). Intelligence estimation from images is a difficult task even for humans, but it is possible to use algorithms to estimate it beyond a random guess. For PI, the correlation is higher ($\rho = 0.36$), since to some extent, the computer is able to extract effective visual features may be used by humans as shown in Figure 4.

The RMSE of computer is lower than that of the Random, but not (much) lower than the Mean. The performance improvement of computer is much better in terms of correlation than in terms of RMSE is because the regressor tends to predict the *rank* between different scores better in terms of Spearman’s coefficient other than the actual *values*. Similar observation has been shown in [21] and this is also compatible with the prevailing opinion on intelligence in psychology that IQ is essentially a rank other than a true unit of intellectual ability [3].

4.3 Discussion

4.3.1 Human bias

Given that the computer’s predictions of a user’s intelligence do not match humans’ perceptions, it begs the question of what features humans use that our image features do not capture. For example, the more complicated high-level semantic features. We manually examined the 50 images with the highest and lowest PI scores (due to the privacy agreement with participants, we are unable to show example images of those two groups). Inevitably this is a subjective exercise but we hope that the following insights will be helpful to guide future algorithms.

The top 50 images which are perceived as high intelligence contain visual cues related to business clothing, books, instruments, chess, science, music, art, university, formal dining, archery, computers, and math. The bottom 50 images contain visual cues such as colourful hair, offensive hand gestures, an overweight body, heavy make-up, smokers, tattoos, and black ethnicity.

Together with the results in Section 4.1, which found that humans use cues, such as wearing glasses, that are uncorrelated with MI, this provides further evidence of how human raters tend to use biased visual cues when judging intelligence from profile images. In contrast, computing approaches may have the potential to reduce human bias.

4.3.2 Application

Profile pictures are a ubiquitous way for users to present themselves on social networks, and typically they are not considered to be private data. Based on our work, an automatic profile picture rating system could be created to determine whether a profile image “looks” intelligent and suitable for professional places such as LinkedIn and on a CV. This can help people to better manage their self-representations. Commercial companies can similarly benefit from such an intelligence rating system by selecting suitable portrait photos in brand representation to make a good impression on customers.

5. CONCLUSION

We have proposed a framework of intelligence estimation from profile images both by humans and computers. Humans are able to make intelligence judgements for both men

and women from their social network profile images. Experimental results show that it is possible to equal humans’ accuracy using computing methods, while also having the potential to reduce biased judgements based on inaccurate stereotypes.

Our results found that people who are measured as intelligent and also perceived as intelligent do not like to use the colour pink, purple or red in their profile images, and their images are usually less diversified in colour, more clear in texture, and contain less skin area. Besides that, intelligent people also like to use the colour green, and have fewer faces in their images, but this does not affect how others judge them. Essentially, the most intelligent people in our dataset understand that a profile picture is most effective when it shows a single person, captured in focus, and with an uncluttered background.

The following cues are inaccurate stereotypes - correlated with perceptions of intelligence but not measured intelligence. Profile images containing more grey and white, but less brown and green, with higher chromatic purity, smiling and wearing glasses, and faces at a proper distance from the camera, make people look intelligent no matter how smart they really are.

Our results show that humans use inaccurate stereotypes to make biased judgements about the perceived intelligence of the person who uses a profile picture, which raises considerable concerns about the common practice for hiring managers to search candidates online before inviting them to interview. Nevertheless, our results also indicate that the choices that users make of how to present themselves in their profile picture are reflective of their measured intelligence, and that computer algorithms can extract features and automatically make predictions.

6. FUTURE WORK

There are several directions for future work. 1) Our results show that high-level features which capture the content and semantic information of images are helpful for intelligence estimation. In the future we will investigate additional high-level features with more sophisticated object detection and image caption techniques. Those interpretable features will benefit the human behaviour research in social science. 2) Psychology research showed that individual differences are influenced by both biological and environmental factors [24]. So it also should consider sociodemographic factors such as ethnicity, language and culture in the analysis for visual features and this will help generate prior knowledge for better feature selection. 3) Other efforts can focus on improving the estimation accuracy such as collecting multiple images for each user then modelling the estimation process using multiple instance regression models or fusing results by multiple classifier system in different levels (e.g., feature, score, rank, etc.).

7. REFERENCES

- [1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041, Dec 2006.
- [2] S. Baert. Do they find you on facebook? facebook profile picture and hiring chances. *IZA Discussion Paper*, 9584, 2015.

- [3] D. J. Bartholomew. *Measuring intelligence: Facts and fallacies*. Cambridge University Press, 2004.
- [4] S. Bhattacharya, B. Nojavanasghari, T. Chen, D. Liu, S.-F. Chang, and M. Shah. Towards a comprehensive computational model for aesthetic assessment of videos. In *ACM Int. Conf. Multimedia (MM)*, pages 361–364, 2013.
- [5] D. V. Cicchetti. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. assessment*, 6(4):284, 1994.
- [6] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina. Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In *ACM Int. Conf. Multimedia (MM)*, pages 213–222, 2013.
- [7] S. C. Guntuku, L. Qiu, S. Roy, W. Lin, and V. Jakhetiya. Do others perceive you as you want them to?: Modeling personality based on selfies. In *Int. Workshop on Affect and Sentiment in Multimedia ASM*, pages 21–26, 2015.
- [8] S. C. Guntuku, J. T. Zhou, S. Roy, L. Weisi, and I. W. Tsang. *Asian Conf. Computer Vision (ACCV)*, chapter Deep Representations to Model User ‘Likes’, pages 3–18. 2015.
- [9] M. C. Keller, C. E. Garver-Apgar, M. J. Wright, N. G. Martin, R. P. Corley, M. C. Stallings, J. K. Hewitt, and B. P. Zietsch. The genetic correlation between height and IQ: Shared genes or assortative mating? *PLoS Genet*, 9(4):1–10, 04 2013.
- [10] K. Kleisner, V. Chvátalová, and J. Flegr. Perceived intelligence is associated with measured intelligence in men but not women. *PLoS ONE*, 9(3):1–7, 03 2014.
- [11] I. Kokkinos. Bounding part scores for rapid detection with deformable part models. In *European Conf. Computer Vision (ECCV)*, pages 41–50, 2012.
- [12] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, 110(15):5802–5805, 2013.
- [13] L. Leqi, D. Preoțiuc-Pietro, Z. Riahi, M. E. Moghaddam, and L. Ungar. Analyzing personality through social media profile picture choice. In *Int. AAAI Conf. Web and Social Media (ICWSM)*, 2016.
- [14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [15] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM Int. Conf. Multimedia (MM)*, pages 83–92, New York, NY, USA, 2010. ACM.
- [16] K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychol. methods*, 1(1):30, 1996.
- [17] J. Raven, J. Raven, and J. Court. Manual for raven’s progressive matrices and vocabulary scales. *San Antonio, TX: Harcourt Assessment*, 2003.
- [18] M. Redi, D. Quercia, L. Graham, and S. Gosling. Like partying? your face says it all. predicting the ambiance of places with profile pictures. In *Int. AAAI Conf. Web and Social Media (ICWSM)*, 2015.
- [19] F. E. Saal, R. G. Downey, and M. A. Lahey. Rating the ratings: Assessing the psychometric quality of rating data. *Psychol. Bull.*, 88(2):413, 1980.
- [20] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, 8(9):1–16, 09 2013.
- [21] 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 Trans. Affect. Comput.*, PP(99):1–1, 2016.
- [22] F. Steele, D. C. Evans, and R. K. Green. Is your profile picture worth 1000 words? photo characteristics associated with personality impression agreement. In *Int. AAAI Conf. Web and Social Media (ICWSM)*, 2009.
- [23] S. N. Talamas, K. I. Mavor, J. Axelsson, T. Sundelin, and D. I. Perrett. Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness. *J. Exp. Psychol.-Gen.*, 145(5):603, 2016.
- [24] S. N. Talamas, K. I. Mavor, and D. I. Perrett. Blinded by beauty: Attractiveness bias and accurate perceptions of academic performance. *PLoS ONE*, 11(2):1–18, 02 2016.
- [25] H. C. Triandis and E. M. Suh. Cultural influences on personality. *Annu. rev. psychol.*, 53(1):133–160, 2002.
- [26] A. Vinciarelli and G. Mohammadi. A survey of personality computing. *IEEE Trans. Affect. Comput.*, 5(3):273–291, July 2014.
- [27] W. Youyou, M. Kosinski, and D. Stillwell. Computer-based personality judgments are more accurate than those made by humans. *PNAS*, 112(4):1036–1040, 2015.
- [28] J. L. Zagorsky. Do you have to be smart to be rich? the impact of IQ on wealth, income and financial distress. *Intelligence*, 35(5):489 – 501, 2007.
- [29] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin. Extensive facial landmark localization with coarse-to-fine convolutional network cascade. In *IEEE Int. Conf. Computer Vision Workshops (ICCVW)*, pages 386–391, Dec. 2013.
- [30] X. Zhu, H. I. Suk, and D. Shen. Matrix-similarity based loss function and feature selection for alzheimer’s disease diagnosis. In *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 3089–3096, June 2014.