Social Impression of Faces: From Prediction to Modification

Amanda Song, Chad Atalla, Bartholomew Tam, Linjie Li, and Garrison Cottrell

Abstract First impressions are influential in shaping our personal, economic, and political decisions. We develop a computational framework that can model and modify impressions of faces. First, we use a state-of-the-art predictive model of facial impressions (such as facial attractiveness, trustworthiness, and intelligence) and apply it to a large-scale natural face dataset in order to create a robust facial impression dataset. We validate the augmented dataset with respect to human judgments. Second, we use the new dataset to train a model, ModifAE, that changes face smoothly and effectively in multiple social dimensions. This modification model offers social scientists the ability to manipulate impressions as needed, and it sheds light on both the biases and the visual features underlying first impression formation.

1 Introduction

Humans quickly form subjective impressions of faces, judging traits like facial attractiveness, trustworthiness, and aggressiveness [1]. Despite the continuous scale and subjective nature of these social judgments, there is often a consensus among humans in how traits are perceived; for example, human raters will agree that certain faces appear relatively more trustworthy [2, 3]. Social judgments of faces have a significant impact on social outcomes, ranging from electoral success to sentencing decisions [4, 5]. Modeling is one way to understand these critical split-second impressions. Another way is through explicit human-judged experiments, which require carefully controlled datasets (e.g., building a dataset of faces that vary in "trustworthiness" while remaining consistent across age, gender, and "attractiveness"). In this work, we develop a system to model these impressions, predict human average impressions on facial images, visualize human perceptual biases, and create isolated image modifications for experimental datasets.

Choosing a subset of social impressions for modeling, we look to the 10k US Adult Faces Database [6]. Bainbridge et al. [6] investigated what social traits influence the

A. Song (✉) · C. Atalla · B. Tam · L. Li · G. Cottrell
University of California, Gilman Dr, La Jolla 9500, San Diego, CA 92122, USA

Z. Yang and E. von Briesen (eds.), *Proceedings of the 2019 International Conference of The Computational Social Science Society of the Americas*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-77517-9_1

memorability of a face. They compiled a list of 20 spontaneous social judgments and the corresponding opposite traits. Then, they assembled a human-judged dataset of trait ratings on 2,222 faces from the 10k US Adult Faces Database. Among the 40 traits, "trustworthy", "attractive,", "aggressive," and "intelligent" were frequently used in human-written face descriptions, played a significant role in face memorability, and had relatively high rating agreement levels between human judges. Therefore, we choose them as a representative subset of social impressions for modeling in this paper. Motivated by the success of deep learning in modeling visual properties, we use deep learning-based pretrained representations as the basis for learning to predict first impressions on realistic faces, training a predictive linear model that successfully predicts human social perception on faces whenever human have consensus.

To create controlled face datasets and visualize perceptual biases, a generative model is needed. Recent generative image models have been successful in creating high-resolution, high-fidelity, and diverse images [7–9]. However, in the face space, most generative models have focused on editing or modifying categorical and objective attributes, such as expression, gender, hair color, and identity [9]. These categorical changes are referred to as "image-to-image translation." Here, we focus on modifying continuous traits of an image, which we refer to as "continuous image modification" [10]. Regarding continuous image modification, there has been work on modifying the memorability [11] and attractiveness of a face [12], but these models do not generalize to wider sets of social impressions. Also, some researchers have endowed computer-generated faces with particular social impressions, but these models cannot modify real face images [5, 13]. So, no prior work has attempted to automatically modify general continuous social impressions of real face photographs. Part of the difficulty lies in the fact that training a high-fidelity generative model requires a large amount of data, yet there is no pre-existing dataset that has tens of thousands of faces with labeled social impression trait scores. We overcome this difficulty by proposing a cost-effective and easy-to-scale-up method to construct a large-scale facial impression dataset.

Conditional generative adversarial networks (GANs) [14] have become the most popular tool for the image-to-image translation task, so we compare against a recent GAN as a state-of-the-art (SOTA) reference point [10, 15, 16]. StarGAN [9] is a SOTA conditional GAN that can modify multiple binary categorical traits of faces at once, maintaining the identity of the face using a "cycle consistency" loss function, which translates the face back to the original one [17]. StarGAN consists of two networks: a generator and a discriminator. The generator takes an image and a set of desired categorical traits, producing a modified image. The discriminator takes an image and makes a prediction about its realism and categorical traits. By comparing the fake images to genuine images, the discriminator gives feedback to the generator about how to make the image and desired traits appear more realistic.

Despite the success of GANs in categorical image-to-image translation, they cannot perform continuous image modification without binarizing the task. GANs typically have many parameters and long training times. They are also sensitive to hyper-parameter selection and the delicate balance between generator and discriminator training. Therefore, they can be difficult to train compared to a single-network

model. Finally, they suffer from a lack of interpretability, offering no means of visualizing or understanding why the model makes the modifications it does.

In this work, we address these architectural concerns while designing a neural network to model and automatically modify continuous-scale face traits (rated from 1 to 9) in real face images. First, we use our deep learning-based predictive linear model [18] to predict human facial impressions of attractiveness [3, 19], trustworthiness [2, 20], aggressiveness [21], and intelligence, to form an augmented dataset. We validated the effectiveness of this dataset augmentation method with human experiments.

With this large-scale realistic facial impression dataset, we train a deep modifying autoencoder, ModifAE, that can smoothly and naturally modify the first impressions of faces. We evaluate the model performance quantitatively and qualitatively and compare it with StarGAN. Notably, our generative model can modify multiple traits at once and can provide visualizations of group average trait features. It is also easy to train. These capabilities make it a powerful tool, which can, for example, modify multiple traits while controlling other high-level attributes, such as gender. We then quantify the actual changes ModifAE makes to modify perceived impressions, shedding light on what geometric features correspond to social impression dimensions.

2 Predicting Social Attributes of Faces

Previous studies have shown that pretrained deep learning models can provide feature representations versatile for related tasks [22]. After comparing multiple off-the-shelf pretrained neural networks, we find that conv5_2 layer of VGG16 (pretrained for object classification) leads to satisfactory results. After obtaining intermediate representations from the pretrained neural networks, we apply Principal Component Analysis to reduce the dimensionality, then train a ridge regression model to produce predictions on each social attribute, respectively. For "trustworthy", "attractive,", "aggressive," and "intelligent", the predictive model's correlations (Spearman rank correlation) with human averaging ratings are 0.73, 0.75, 0.72 , and 0.62, respectively, on the test set.

3 Creating a Large-Scale Facial Impression Dataset

To train a generative model on continuous face traits, we need a large and diverse dataset. We use images from the CelebA dataset [23], which consists of over 190,000 images of celebrities. The images in CelebA are annotated with binary categorical labels such as "wearing a hat" but not on continuous ratings of social impressions.

To generate continuous social impression traits of these faces, we use our predictive model mentioned above [18] trained on a smaller dataset (approximately 2,000

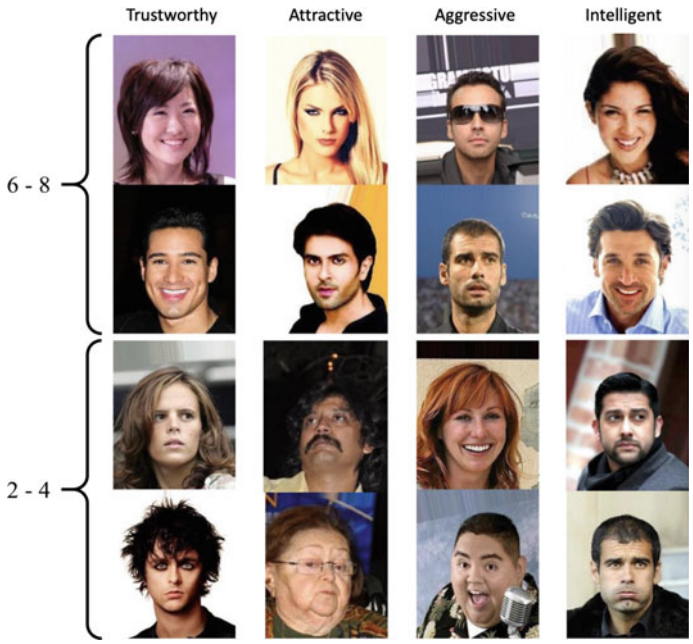

Fig. 1 CelebA faces and their predicted trait scores

faces from the 10k US Adult Faces Database [6]) that have been annotated with continuous ratings by 15 raters for each of the social traits.

We apply this model to over 190,000 faces from the CelebA dataset. The resulting model predictions are highly correlated with human judgments (*** denotes that $p < 0.0001$): trustworthy (0.73***), attractive (0.90***), aggressive (0.76***), and intelligent (0.62***). These correlations are obtained by asking subject to rate faces on these four traits and computing average human ratings' Spearman correlation with the model's predictions. The details of these experiments are given in the Methods.

Example faces and their predicted ratings are shown in Fig. 1. Note here that 6–8 are high ratings, and 2–4 are low ratings.

4 Validating the Algorithm-Augmented Dataset

To evaluate the effectiveness of this algorithm-augmented dataset, we collect human judgments of faces in CelebA and evaluate how model predictions correlate with human judgments. We examined four traits: attractive, aggressive, trustworthy and intelligent because they represent different aspects of first impressions and are of relatively high human agreement. For each trait, we chose 80 faces whose predicted scores are evenly spread across a range of predictions (i.e., from 2 to 8). Every

participant is presented with a random sequence of these 80 faces, and is asked to give each face a rating on a 1–9 scale for the specified trait. Every face is rated by roughly 15 subjects (ranging from 12 to 16), and we compute the average ratings for each face. Lastly, we compute the Spearman rank correlation between the average human ratings and the model's predictions. For all four traits, human average ratings are significantly correlated with model predictions (*** indicates $p < 0.001$): trustworthy (0.73***), attractive (0.90***), aggressive (0.76***), and intelligent (0.62***). We plan to publicly release the large-scale facial impression dataset for future researchers' use.

5 ModifAE: A Modification Model of Social Impressions

With the large-scale validated first impression face dataset, we train our modification model, ModifAE. The network is trained on an autoencoding task (reproducing the input on the output) with an added input corresponding to the trait value (see Fig. 2). By using aggressive dropout on the image side (a technique where a random half of the activations are set to 0), the model implicitly learns to depend on the input trait value to generate the reconstruction of the image. This enables the model, after training, to use different trait values to modify the image.

5.1 Architecture

The ModifAE architecture consists of a single autoencoder with two (image and trait) sets of inputs that pass through an encoding stage and then are fused (by averaging them) in the middle of the network. This latent representation is then fed into an image decoder.

The image encoder and decoder are identical to the encode and decode portions of the StarGAN generator network, scaled to fewer channels [9]. More specifically, the network has two downsampling convolutional layers with stride two, four residual blocks, a bottleneck with 16 channels, four more residual blocks, then two upsampling transposed convolutional layers with stride two [9]. All layers have ReLU activation. We use the first half of this network (including the bottleneck) as the image encoder. We use the remainder of the network as the image decoder. Theoretically, this portion could consist of the encode and decode halves of any other image autoencoder.

The trait encoder takes a one-dimensional set of traits, feeds these into a single dense layer with Leaky ReLU activation, and reshapes the output to create a vector of the identical shape as the image encoder output. The outputs of the trait and image encoders are then combined into a single latent representation of the image and ratings.

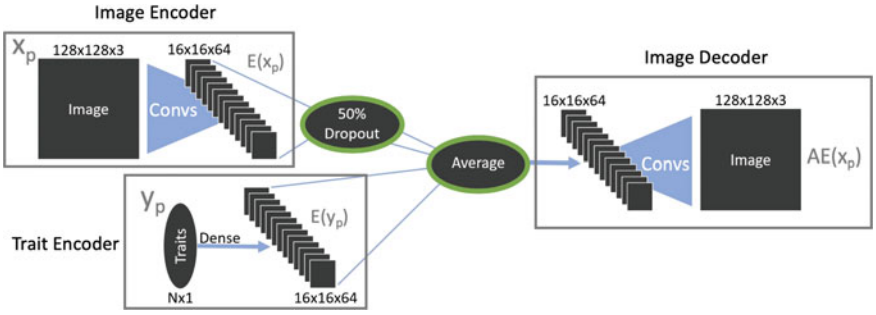

Fig. 2 General illustration of ModifAE architecture

In order to encourage the model to encode the trait information, which is otherwise unnecessary to reproduce the image, 50% dropout is applied to the values from the image encoder. This is then averaged with the trait encoder output to arrive at the combined latent representation. The image decoder projects the representation back into image space, creating the single output image. The architecture is depicted in Fig. 2.

Despite sharing some aspects of architecture with StarGAN's generator [9], ModifAE has over 50 times fewer parameters.

5.2 ModifAE Training Procedure

ModifAE is only trained on an autoencoding task. We train ModifAE using the Adam optimizer [24] and train for 100 epochs on our augmented CelebA images [23]. The objective is to optimize a single loss function based on two terms. We use the L_1 loss on the image autoencoder. We also optimize the L_1 loss between the trait encoder and image encoder. The total loss is

$$L = \frac{1}{N}\sum_{p=1}^{N} |x_p - AE(x_p)| + |E(x_p) - E(y_p)| \tag{1}$$

where x_p is the pth image example, y_p is its trait vector, $E(\cdot)$ is the result of the trait or image encoder, and $AE(\cdot)$ is the output of the full-architecture autoencoder. The second term in this loss function encourages the network to have a similar representation between the trait and the image encodings.

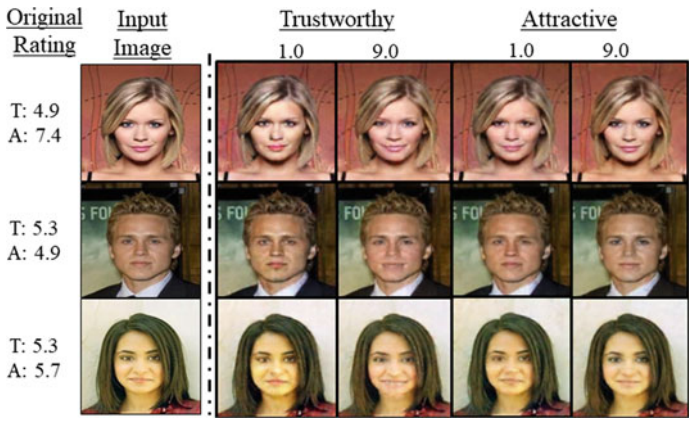

Fig. 3 Image modifications by ModifAE in trustworthy and attractive

5.3 How ModifAE Works

Each image is encoded along with its predicted traits. The image encoder compresses the image down to a bottlenecked latent space, where higher level features about the image are encoded. Simultaneously, the trait encoder projects the given traits to the same latent space, creating an average face representation with those ratings (Fig. 3).

Results

First, we qualitatively examine ModifAE's modifications and visualizations of trait representations and then quantitatively compare the modifications of ModifAE and StarGAN with human behavioral studies.

5.4 Qualitative Evaluation

We obtain visualizations of ModifAE's trait representations by presenting the model with trait values in the absence of any image input. The resulting transformation maps show ModifAE's representation of a trait at different trait values. These transformation maps can be produced from models which were trained on multiple traits, enabling visualizations of how ModifAE perceives some traits to vary independently of others. Figure 5 shows traversals of"attractive," "intelligent," "trustworthy," and "aggressive," while holding gender constant. Through this method, ModifAE addresses the issue of interpretability in generative models. These images provide a

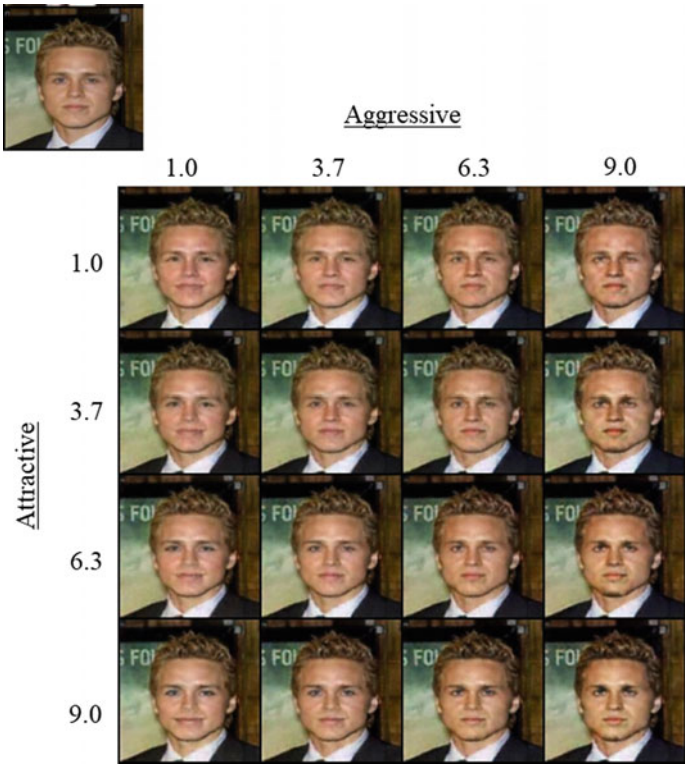

Fig. 4 Multi-trait image modification by ModifAE

window into how ModifAE represents each trait and how ModifAE changes a face to increase or decrease a given trait value.

In general, it appears that faces that subtend a larger visual angle are rated more positively, with the trend for the faces to get bigger from left to right for the three positive traits, and the opposite for aggressive. To our knowledge, this has not been observed previously and hence is a prediction of our model.

Similarly, a larger smile results in more positive ratings, with big smiles on the right for the positive traits and on the left for the negative trait. This accords with our intuition and is consistent with previous research that demonstrates smiling is associated with positive person perception [25].

For attractiveness, in addition to the larger smile corresponding to more attractive, at the unattractive end of the scale, there is lower contrast in the face features.

We also are able to modify two traits at once, by training on both trait values in a single network (see Fig. 4). For this experiment, we trained ModifAE on two traits: "attractive" and "aggressive." The picture in the upper left corner is the original. Looking at the (1, 1) point in Fig. 4 (unattractive and not aggressive), the man's mouth is fairly neutral, and his features are not very pronounced. As attractiveness

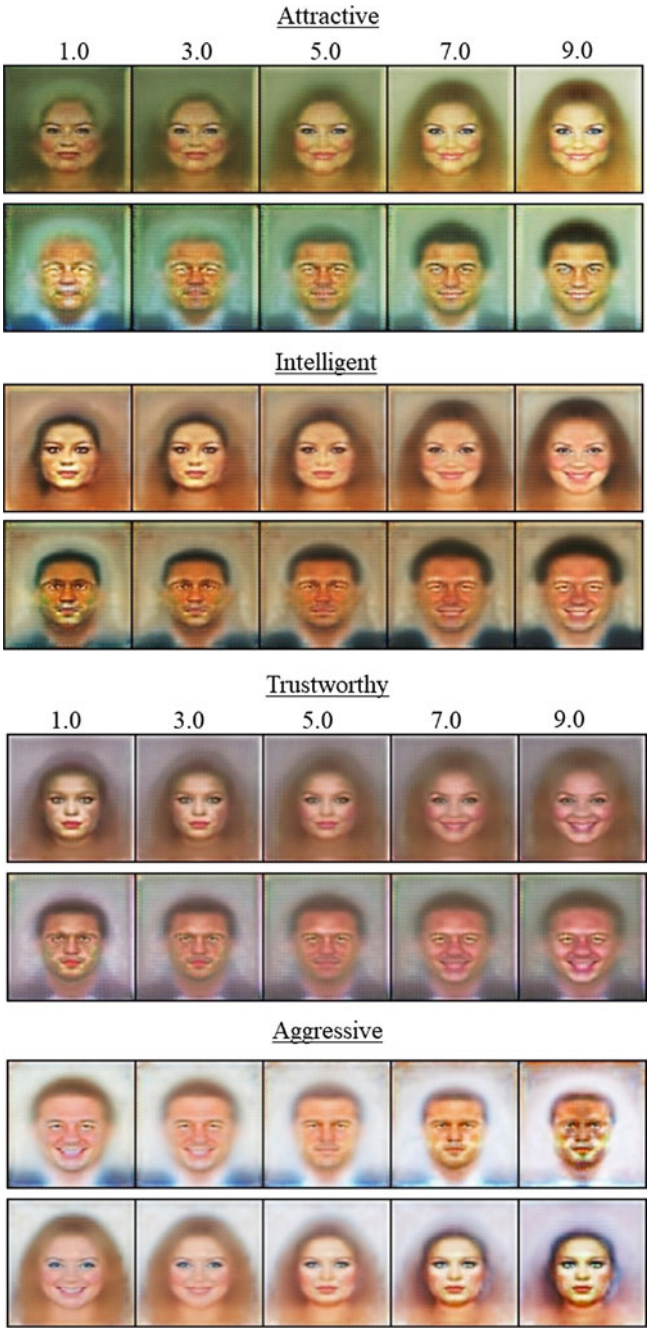

Fig. 5 Visualization of model's internal perception of traits. Each is a traversal of a trait (increasing left to right) while gender is held constant

and aggressiveness increase, the angles of the face become sharper, there is more definition of features like eyes and eyebrows, and the smile shrinks.

In the "attractive" transformation map of Fig. 5, age is a salient factor, with rounder, pudgier, older faces appearing on the left side of the traversal, and faces with sharper features, clearly defined large eyes, and larger smiles appear on the more attractive side. This makes sense and is a confirmation that the model reflects human biases.

In the "intelligent" transformation maps, the degree of smile increases left to right, the hair gets longer, and the eyebrows get smaller, perhaps suggesting that large brows are perceived as less intelligent. Additionally, the head size clearly increases from left to right. This suggests that images in which the head subtends a greater visual angle are perceived as more intelligent, a bias that, to our knowledge, has not been previously observed. Of course, in this case, it is not the real-world size that matters, just the size of the head in the image. This is a prediction of our model.

Similar to the attractiveness traversal, the trustworthiness increases with the degree of smiling. The forehead gets larger, and there is a slight lightening of the hair, suggesting older people are more trustworthy.

For aggressiveness, clearly the bigger the smile, the less aggressive the face appears, which fits well with our intuition. Also, the visual angle of the face becomes smaller, and to us, at least, the eyes appear "beadier." Unfortunately, there also appears to be a racial bias, with darker skin appearing more aggressive.

5.5 Quantitative Evaluation

To evaluate the quality of ModifAE's continuous subjective trait modifications, we performed Amazon Mechanical Turk (AMT) experiments on the four traits we focus on in this article: aggressive, attractive, trustworthy, and intelligent. For each trait, we created 90 image pairs, of which 80 are the same identity modified to be at high and low values of each trait. For StarGAN, we used a median split of low and high-rated traits to train the model, making the transformation binary. ModifAE was trained as previously described. For each model, then, faces were modified to be low or high on each trait. Subjects judged which face had more of the particular trait. 10 pairs were repeats in order to judge subject consistency, and 10 pairs were unmodified CelebA faces with high and low ratings. This latter we called "ground truth" pairs to test whether subjects were paying attention. Subjects whose ratings on these pairs were at chance or below were rejected.

Hence, for each trait, we present participants with a sequence of 100 image pairs, and participants are asked to pick which image most exemplifies the trait in each pair.[1] Each pair was evaluated by 15 subjects.

[1] In a pilot experiment, we asked subjects to rate faces with different identities generated in a fine continuum, but found significant variance with no correlation to the intended scores, presumably because the images were not differentiable at that fine a grain.

Table 1 Comparison of ModifAE with StarGAN

Attribute	ModifAE	StarGAN	"Ground Truth"
Aggressive	0.68***	0.72***	0.90***
Attractive	0.68***	0.51	0.94 ***
Trustworthy	0.63***	0.40	0.87***
Intelligent	0.68***	0.58***	0.81***

*$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$

We calculate the fraction of pairs in which subjects chose the image with the higher modified trait across all participants and all pairs. If they choose the face that was modified to be higher in the trait, then they agree with the model's modifications. The results are shown in Table 1. We perform a binomial test to determine whether each trait's accuracy is significantly below or above chance (***$p < 0.001$). Note that the fourth column "Ground Truth" indicates the overall accuracy of the unmodified "ground truth" pairs. Given the variance in human impression judgments, these numbers serve as a reference ceiling for how well the models can perform.

From Table 1, we can see that for all four traits, ModifAE produces pairs that yield above chance level human agreement. In three out of the four traits, ModifAE significantly outperforms StarGAN; whereas, for the aggressive trait, StarGAN performs only slightly better than ModifAE. StarGAN is good at creating discrete changes in facial expressions, which accounts for this advantage.

Since ModifAE is able to generate continuous modifications, we evaluated this property by creating two more same-face pairs: ones modified to have low values and middle values, and ones modified to have middle values and high values. We obtain human agreement (accuracy) over the Low-Mid and Mid-High pairs for each of the four traits. The results are shown in Table 2.

From Table 2, we find that all the low-mid pairs yield significantly above chance accuracy, yet for mid-high level, only trustworthy pairs have accuracy slightly above chance ($p < 0.05$*). This suggests that human psychological face space is nonlinear and has more differentiation toward the low- to mid-range of social dimensions. Another possibility is that when our model generates faces that are of more extreme scores (e.g., 8 or 9), the model is extrapolating and produces artifacts that lead to that face being rejected. This speculation requires further analysis to be confirmed.

Table 2 ModifAE Low-Mid-High level self-comparison

Attribute	Low-Mid	Mid-High	Low-High
Aggressive	0.60***	0.52	0.68***
Attractive	0.59***	0.52	0.68***
Trustworthy	0.61***	0.53*	0.63***
Intelligent	0.60***	0.50	0.68***

*$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$

5.6 Qualitative Interpretations

With a hypothesis-driven approach, psychologists have identified certain visual features that contribute to specific impressions. The symmetry of the face [26] can explain why certain faces look more attractive. Other global face features such as femininity, babyfaceness [27], typicality [28], and facial width-to-height ratio (fWHR) [29] drive different aspects of social impression perception (warmth, honesty, submissiveness, dominance, etc.). Emotions such as perceived anger and happiness drive aggressiveness and trustworthiness perceptions, respectively. Using morphing and averaging methods, studies [30] have established that age also serves an important role in social perception of attractiveness, trustworthiness, and dominance.

6 Discussion

We have shown that a deep network can be used to predict the human social perception of faces, achieving a high correlation with the average human ratings. As far as we know, this is the widest exploration of social judgment predictions, showing human-like perceptions on 40 social dimensions. By predicting this as a continuous value, rather than categorical, the subjective nature of human judgment is modeled smoothly, along with the subjective face trait landscape.

Of greater significance is our model's correlations with human judgments for traits such as trustworthiness, responsibility, confidence, and intelligence, which correspond to more static features of the face. In this area, the deep network, which responds to facial textures and shape, has superior performance. While these judgments do not correspond to the traditional notion of "ground truth", they are descriptions for which humans have a fair amount of agreement, suggesting the presence of a signal to be recognized. Furthermore, we have shown that our prediction model can generalize reasonably well to an entirely new dataset, making it widely applicable to real-world scenarios.

We further develop a generative model, ModifAE, which can modify a face's social impressions while preserving its realism. ModifAE can change a face's perceived social features (e.g., make a face look more sociable, trustworthy). It can also produce transformation maps that elegantly summarize the average opinions and biases of a group of raters who have created a dataset. This functionality enables psychologists to quantify human biases during the formation of social impressions in a precise and systematic manner. Psychologists could generate variants of a real face differing in age, gender, and race while holding other traits constant. This controlled dataset could be used to explore how various factors separately and jointly affect the social impressions of faces.

Our computational models make predictions and modifications regarding the first impressions of faces, and such first impressions are indicative of implicit bias toward different social groups [31]. With knowledge of people's first impressions, along

with the embedded potential bias, we have a chance to analyze the perceptual and social interactions that are fundamental to humans.

These results are also significant for the field of social robotics and the fight against discrimination. Predictive models like this can bring empathy to robotics, where technology can help us bridge the emotional and social divide and promote social equality. Empathetic technology can benefit people who are implicitly discriminated against based on social impressions. While a robot should not purely judge a human on appearance, much of human interaction is dictated by the underlying fabric of social impressions. Thus, it is important for a robot to be aware of the subjective social fabric, opening the door to useful knowledge such as whether humans might judge a person to be trustworthy. These judgments may happen subconsciously for humans, while a robot can be more objective, predicting these judgments and objectively choosing when to consider them in a decision. A robot need not treat an attractive or unattractive person differently for its own purposes, but this knowledge could affect how interactions are made for the sake of the human, knowing in advance how that person may feel that they fit into the social landscape. These applications can have significant societal effects.

References

1. Todorov, A., Olivola, C.Y., Dotsch, R., Mende-Siedlecki, P.: Social attributions from faces: determinants, consequences, accuracy, and functional significance. Annu. Rev. Psychol. **66**(1), 519 (2015)
2. Falvello, V., Vinson, M., Ferrari, C., Todorov, A.: The robustness of learning about the trustworthiness of other people. Soc. Cogn. **33**(5), 368 (2015)
3. Eisenthal, Y., Dror, G., Ruppin, E.: Facial attractiveness: beauty and the machine. Neural Comput. **18**(1), 119–142 (2006)
4. Dumas, R., Testé, B.: The influence of criminal facial stereotypes on juridic judgments. Swiss J. Psychol. **65**(4), 237–244 (2006)
5. Oosterhof, N.N., Todorov, A.: The functional basis of face evaluation. Proc. Natl. Acad. Sci. **105**(32), 11087–11092 (2008)
6. Bainbridge, W.A., Isola, P., Oliva, A.: The intrinsic memorability of face photographs. J. Exp. Psychol.: Gen. **142**(4), 1323 (2013)
7. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis (2018). arXiv:1809.11096
8. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation (2017). arXiv:1710.10196
9. Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S., Choo, J.: Stargan: unified generative adversarial networks for multi-domain image-to-image translation. CoRR (2017) arXiv:1711.09020
10. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CoRR (2016). arXiv:1611.07004
11. Khosla, A., Bainbridge, W.A., Torralba, A., Oliva, A.: Modifying the memorability of face photographs. In: International Conference on Computer Vision (ICCV-2013), pp. 3200–3207. IEEE (2013)
12. Leyvand, T., Cohen-Or, D., Dror, G., Lischinski, D.: Data-driven enhancement of facial attractiveness. ACM Trans. Graph. (TOG) **27**(3), 38 (2008)
13. Vernon, R.J., Sutherland, C.A., Young, A.W., Hartley, T.: Modeling first impressions from highly variable facial images. Proc. Natl. Acad. Sci. **111**(32), E3353–E3361 (2014)

14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
15. Mirza M., Osindero, S.: Conditional generative adversarial nets. CoRR (2014). arXiv:1411.1784
16. Lee, M., Seok, J.: Controllable generative adversarial network. CoRR (2017). arXiv:1708.00598
17. Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
18. Song, A., Li, L., Atalla, C., Cottrell, G.: Learning to see people like people (2017). arXiv:1705.04282
19. Gray, D., Yu, K., Xu, W., Gong, Y.: Predicting facial beauty without landmarks. In: Computer Vision–ECCV 2010, pp. 434–447. Springer, Berlin (2010)
20. Todorov, A., Baron, S.G., Oosterhof, N.N.: Evaluating face trustworthiness: a model based approach. Soc. Cogn. Affect. Neurosci. **3**(2), 119–127 (2008)
21. Mignault, A., Chaudhuri, A.: The many faces of a neutral face: head tilt and perception of dominance and emotion. J. Nonverbal Behav. **27**(2), 111–132 (2003)
22. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
23. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (2015)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR (2014). arXiv:1412.6980
25. Reis, H.T., Wilson, I.M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., Franco, M., Gioioso, E., Freeman, L., Radoane, K.: What is smiling is beautiful and good. Eur. J. Soc. Psychol. **20**(3), 259–267 (1990)
26. Scheib, J.E., Gangestad, S.W., Thornhill, R.: Facial attractiveness, symmetry and cues of good genes. Proc. R. Soc. London. Ser. B: Biol. Sci. **266**(1431), 1913–1917 (1999)
27. Berry, D.S., Zebrowitz-McArthur, L.: What's in a face? facial maturity and the attribution of legal responsibility. Pers. Soc. Psychol. Bull. **14**(1), 23–33 (1988)
28. Sofer, C., Dotsch, R., Wigboldus, D.H., Todorov, A.: What is typical is good: the influence of face typicality on perceived trustworthiness. Psychol. Sci. **26**(1), 39–47 (2015)
29. Haselhuhn, M.P., Ormiston, M.E., Wong, E.M.: Men's facial width-to-height ratio predicts aggression: a meta-analysis. PLoS One **10**(4), (2015)
30. Sutherland, C.A., Oldmeadow, J.A., Santos, I.M., Towler, J., Burt, D.M., Young, A.W.: Social inferences from faces: ambient images generate a three-dimensional model. Cognition **127**(1), 105–118 (2013)
31. Stanley, D.A., Sokol-Hessner, P., Banaji, M.R., Phelps, E.A.: Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc. Natl. Acad. Sci. **108**(19), 7710–7715 (2011)