How does beauty lie in the eye of the beholder?

Anonymous Author(s)
Affiliation
Address
email

Abstract

Is beauty in the eye of the beholder? To answer this question, we explore an attractiveness rating dataset which contains 1500 raters' rating to the attractiveness level of 200 faces. We examine the commonalities and differences in people's tastes for attractiveness, and we explore what features makes a face attractiveness. We employ different facial features to predict the population attractiveness ratings and analyze what features contribute to attractiveness, and to what extent people share consensus regarding the importance of those features. More interestingly about the dataset is that the 1500 raters are 750 pairs of twins. So this gives us a chance to examine whether people who share more genes also share more in their aesthetic taste.

1 Introduction

What makes a face beautiful has been studied by previous researchers in psychology, machine learning and philosophy field. Psychologists usually examine one factor at a time. For instance, they have studied the role that symmetry and averageness play in attractiveness perception and they found that in general symmetric faces are more attractive but the most attractive faces are asymmetric. Machine learning researchers have worked on automatically predicting the facial attractiveness using computer vision features or deep neural network. But less attention is paid on the interpretability of the visual features that make good predictions. Moreover, there is little work on how people's preference differ from each other and whether there are sub-populations sharing similar tastes on attractiveness. Philosophers care about whether beauty is in the face itself or in the eye of the beholder. Experimental psychologists conduct studies to address this question. In a study in [5], the authors recruited 750 pairs of twins and ask them to rate the same set of 200 faces and they then use statistical analysis to examine whether people who share more genes will share more in their attractiveness tastes. And their answer is no. But they didn't explicitly examine what features makes a face attractive, and how different people weigh on those features. Our research aims at bridging this gap. We chose a set of interpretable facial features and use them to predict facial attractiveness from different raters. We then examine how people weigh various features differently.

2 Related Work

Most machine learning work on attractiveness prediction focus on predicting the universal attractiveness: attractiveness rating of a face is averaged across a large population as the ground truth to train the algorithms. In [1], Aarabi, et al adopt kNN to classify a face as one of the four rating of attractiveness. They chose 8 geometric ratios of distances between certain fiducial points of the face. On a validation set of 40 images, they achieved 91% accuracy in classification.

White, et al

3 Datasets

In [5], the researcher recruit 547 MZ twin pairs and 214 same-sex DZ twin pairs. Every participant is asked to give a subjective rating to 200 faces from 1-7 scale (1 stands for least attractive whereas 7 for most attractive). The 200 faces are taken from four sources: (1) The MIT face database[3], (2) the Glasgow Unfamiliar Face Database (GUFD)[15], (3) GenHead software, and (4) several databases. Details of the database selection can be found in [5].

Some example faces from the four datasets are shown below.

4 Population's taste: general analysis

People share a certain degree of consensus regarding what is an attractive face.

To quantatively evaluate how much consensus people share regarding their criteria for attractiveness, we compute three main correlation metrics: Pearson correlation, Spearman Rank Correlation and Kendall's Tau.

Kendall'tau is a measure of the correspondence between two ranking. Values close to 1 indicate strong agreement, values close to -1 indicate strong disagreement. This is the tau-b version of Kendall's tau which accounts for ties.

5 Visual Feature Selection

When extracting visual features, we have a number of choices. Traditionally, researchers find that geometric features(the ratio between different parts of the faces and the size of eyes, nose, and etc.) are informative about facial attractiveness. Averageness and symmetry are also two frequently studied features favored by psychologists[7, 12, 11]. In traditional computer vision, eigenface features[13] and Gabor filter features[6], HoG[4], SIFT features[14] and etc. are well studied and are primitive features extractors of faces.

In the last few years, the rapid development of deep learning have swept machine learning and computer vision field. In computer vision, convolutional neural networks (CNNs) constantly refresh records in object recognition, object detection, semantic segmentation and various tasks. These progresses demonstrate that the representational capacity of CNNs has surpassed the old-fashioned hand-crafted visual features. More excitingly, many studies have shown that the visual features learned by convolutional neural networks for image recognition tasks are transferable to facilitate other tasks [16]. Therefore, CNN is also served as an effective visual feature extractor, providing high level representation of a visual input.

5.1 Geometric Features

Ratio in the face matters. People have noticed that golden ratio [10, 9]. We take a set of 29 facial features related to the configurations and ratios between different parts of the faces as our initial set of features. See Table 1. These features are borrowed from [2] and [8]. We also add symmetry and averageness into it. (To be done.)

5.2 Geometric Feature List

5.2.1 Feature Extraction

We first extract 68 face landmarks by dlib, a C++ library and then use the landmarks to compute the facial features listed in Table 1. The 68 face landmarks are shown in figure1.

5.2.2 PCA Analysis of the configurational features

For the 200 faces, we first computed the 29 features, then conduct Principal Component Analysis on the 200 $\times$ 29 feature matrix to analyze the correlation and dependency among features.

Facial Feature	Measurement
Nose Width	Distance between outside edge of the nose at widest point
Nose Length	Distance between nose tip and upper edge of eyes at nose tip center
Lip Thickness	Distance between top and bottom edge of lips at thickest point
Face Length	Distance between bottom of chin to edge of top of forehead/ hairline
Eye Height	Distance between upper and lower inner eyelid at pupil center (Right and left measured separately and then averaged)
Eye Width	Distance between inner and outer corner of eye (Right and left measured separately and then averaged)
Face Width at Most Prominent Part of the Cheek	Distance between the outer edges of cheeks at mid-mouth
Face Width at Mouth	Distance between outer edges of cheeks at mid-mouth
Forehead Length	Distance from center of top of forehead/hairline to the center between the ey
Distance between Pupils	Distance between the center of each pupil
Distance between Pupil and Top of face	Distance between pupil center to top of forehead/hairline (Right and left measured separately and then averaged)
Distance between Pupil and Upper Lip	Distance between pupil center to top edge of lips (Right and left measured separately and then averaged)
Chin Length	Distance from the bottom edge of lips to base of chin
Length of Cheek to Chin	Distance between mid-cheek to bottom of chin (Right and left measured separately and then averaged)
Mid-brow to Hairline	Distance between middle eyebrow to top of forehead/hairline (Right and left measured separately and then averaged)
Facial Width-to-Height Ratio (fWHR)	(Distance between the outer edges of the cheek at most prominent point) ÷ (Distance between upper lip and brow)
Face shape	(Face Width at Most Prominent Part of the Cheek) ÷ (Face Length)
Heartshapeness	(Face Width at Most Prominent Part of the Cheek) ÷ (Face Width at Mouth
Nose shape	(Nose Width) ÷ (Nose Length)
Lip fullness	(Lip Thickness) ÷ (Face Length)
Eye shape	(Eye Height) ÷ (Eye Width)
Eye size	(Eye Height) ÷ (Face Length)
Upper Head Length	(Forehead length) ÷ (Face Length)
Midface Length	(Distance between pupil and upper lip averaged for right and left side) ÷ (F
Chin size	(Chin Length) ÷ (Face Length)
Forehead Height	(Mid-brow to Hairline averaged for right and left side) ÷ (Face Length)
Cheekbone Height	(Length of cheek to chin averaged for right and left side) ÷ (Face Length)
Cheekbone Prominence	(Face width at most prominent part of the cheek) ÷ (Face Length)
Face Roundness	(Face width at mouth) ÷ (Face Length)

Table 1: Facial Feature List

Figure 1: 68 face landmarks labeled by dlib software automatically

We first compute how many PC components can capture 99% of the variance and find that the first eight principal components can capture 99% of the variance and show the distribution in figure2.

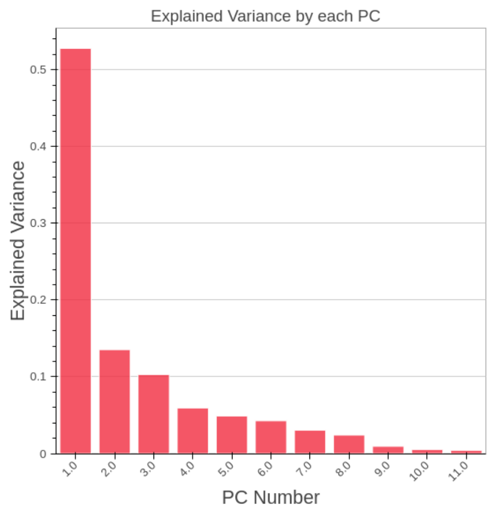

Figure 2: Distribution of explained variance

We then sort the 200 images in each of the first eight PC in ascending sequence, and visualize the images.

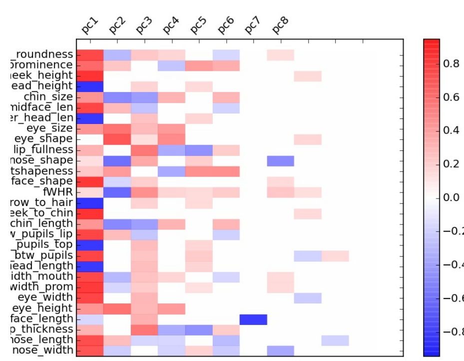

Figure 3: How Principal Components Correlate with the original features

6 Linear Regression Analysis on attractiveness prediction

6.1 Basic Linear Regression

We first apply the very basic linear regression models. We use the eight PC components which retain 95% of the feature variance as the image features and fit a separate model for every subject. For every subject, there are 200 data points, and we conduct a 10-fold split, using 90% of the data as training data and 10% of the data as testing data. We report the correlation coefficient, MSE(Mean Squared Error), MAE(Mean Absolute Error), r2 scorea measure of goodness-of-fit, range from 0 to 1, the larger the better) on training and testing set4.

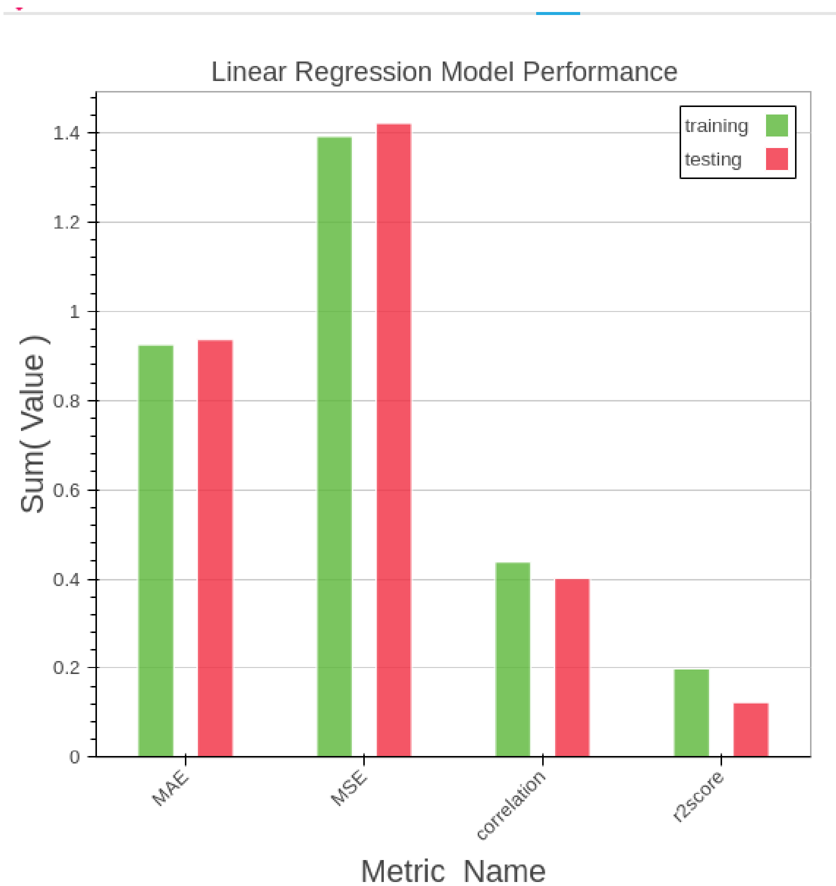

Figure 4: Linear Regression Performance on training and testing set: 10 fold split

6.2 Ridge Regression

Next, we apply the Lasso Regression on the attractiveness rating data. We feed the original features into the ridge regression. Ridge regression is linear regression with L-2 norm regularization so it will encourage smaller coefficients.

6.3 Lasso Regression

We apply Lasso Regression on the dataset and feed the original visual features into the algorithm. Lasso is linear regression with L-1 norm regularizations and it will make most of the coefficients equal to zero.

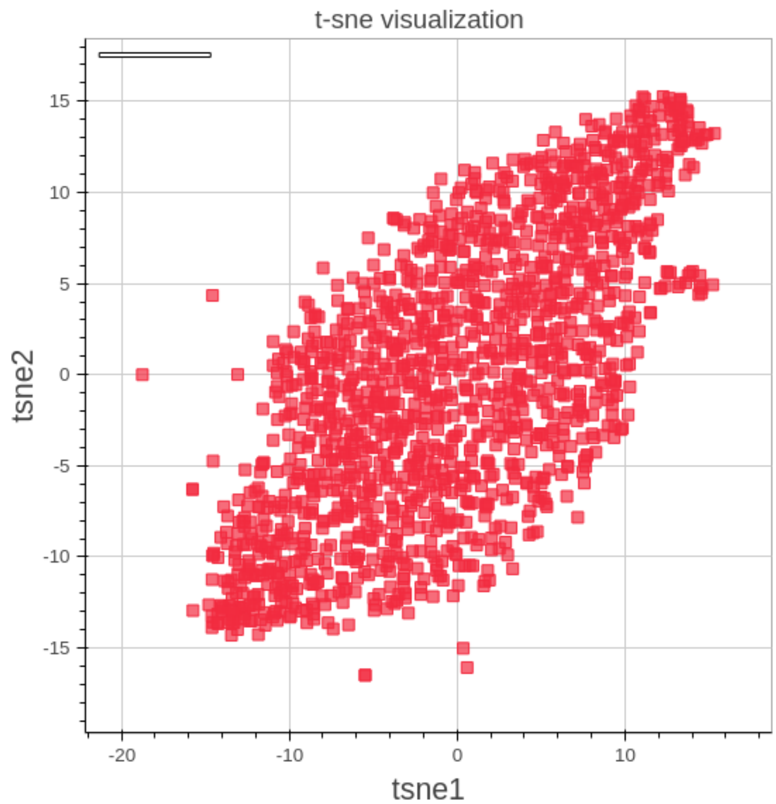

Figure 5: 2D visualization on the original rating matrix

6.4 Linear Rank correlation

6.5 Regression Tree

6.6 Model comparison

6.7 Statistical test

7 SVM Regression Analysis on attractiveness prediction

8 Population Analysis

8.1 tSNE of raters' original ratings

We first use tSNE to conduct dimensionality reduction on the original rating matrix. After dimensionality reduction, the visualization is as follows. We can see that there is no salient clusters. 5

We use the output of tSNE as a compact representation of each rater, so each rater is represented by a 2 dimensional vector (tsne1 and tsne 2 dimension values).

8.2 tSNE of raters' taste vectors

We then use the coefficients computed from linear regression to represent the taste for every rater and get a 1540 * 9 (num-rater by num-feature) matrix and feed it to tsne.

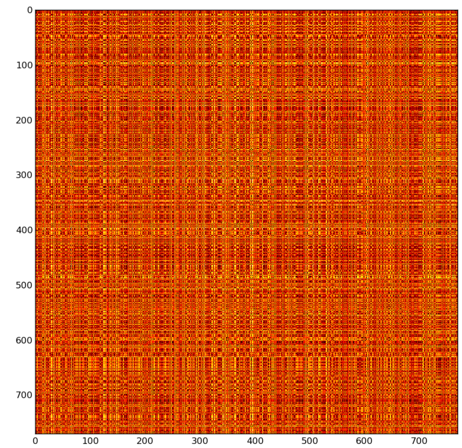

Figure 6: HeatmapOnPairwiseDistances

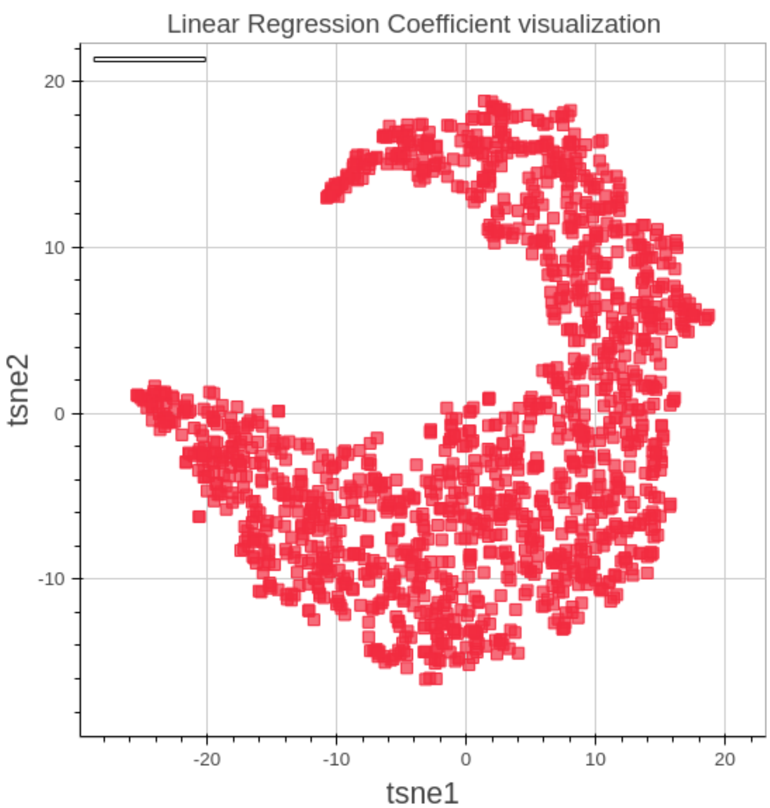

Figure 7: 2D visualization on the original rating matrix

9 Twin vs non-twin analysis

10 Other visual features

10.1 EigenFace features

10.2 VGG face features

10.3 VGG object features

10.4 Alexnet features

10.5 GoogLeNet features

10.6 Visualize what each neural net features encode for

11 Conclusions

12 Discussion

References

[1] P. Aarabi, D. Hughes, K. Mohajer, and M. Emami. The automatic measurement of facial beauty. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 4, pages 2644–2647. IEEE, 2001.

[2] I. V. Blair and C. M. Judd. Afrocentric facial features and stereotyping. *The science of social vision*, 18:306–320, 2011.

[3] P. M. Bronstad and R. Russell. Beauty is in the weof the beholder: Greater agreement on facial attractiveness among close relations. *Perception*, 36(11):1674–1681, 2007.

[4] O. Déniz, G. Bueno, J. Salido, and F. De la Torre. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 32(12):1598–1603, 2011.

[5] L. Germine, R. Russell, P. M. Bronstad, G. A. Blokland, J. W. Smoller, H. Kwok, S. E. Anthony, K. Nakayama, G. Rhodes, and J. B. Wilmer. Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology*, 25(20):2684–2689, 2015.

[6] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. In *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, pages 14–19. IEEE, 1990.

[7] J. H. Langlois and L. A. Roggman. Attractive faces are only average. *Psychological science*, 1(2):115–121, 1990.

[8] D. S. Ma, J. Correll, and B. Wittenbrink. The chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, 47(4):1122–1135, 2015.

[9] A. Nikgoo, K. Alavi, K. Alavi, and A. Mirfazaelian. Assessment of the golden ratio in pleasing smiles. *World journal of orthodontics*, 10(3), 2009.

[10] P. M. Pallett, S. Link, and K. Lee. New golden ratios for facial beauty. *Vision research*, 50(2):149–154, 2010.

[11] D. I. Perrett, D. M. Burt, I. S. Penton-Voak, K. J. Lee, D. A. Rowland, and R. Edwards. Symmetry and human facial attractiveness. *Evolution and human behavior*, 20(5):295–307, 1999.

[12] G. Rhodes, F. Proffitt, J. M. Grady, and A. Sumich. Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, 5(4):659–669, 1998.

[13] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[14] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1469–1472. ACM, 2010.

[15] E. A. Vessel and N. Rubin. Beauty and the beholder: highly individual taste for abstract, but not real-world images. *Journal of vision*, 10(2):18–18, 2010.

[16] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, pages 3320–3328, 2014.