# SCUT-FBP5500‡: A Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction

Lingyu Liang, Luojun Lin, Lianwen Jin*, Duorui Xie and Mengru Li
South China University of Technology, Guangzhou 510641, China
E-mail:{lianwen.jin, lianglysky}@gmail.com
‡Dataset Download URL: https://github.com/HCIILAB/SCUT-FBP5500-Database-Release

*Abstract*—Facial beauty prediction (FBP) is a significant visual recognition problem to make assessment of facial attractiveness that is consistent to human perception. To tackle this problem, various data-driven models, especially state-of-the-art deep learning techniques, were introduced, and benchmark dataset become one of the essential elements to achieve FBP. Previous works have formulated the recognition of facial beauty as a specific supervised learning problem of classification, regression or ranking, which indicates that FBP is intrinsically a computation problem with multiple paradigms. However, most of FBP benchmark datasets were built under specific computation constrains, which limits the performance and flexibility of the computational model trained on the dataset. In this paper, we argue that FBP is a multi-paradigm computation problem, and propose a new diverse benchmark dataset, called SCUT-FBP5500, to achieve multi-paradigm facial beauty prediction. The SCUT-FBP5500 dataset has totally 5500 frontal faces with diverse properties (male/female, Asian/Caucasian, ages) and diverse labels (face landmarks, beauty scores within [1, 5], beauty score distribution), which allows different computational models with different FBP paradigms, such as appearance-based/shape-based facial beauty classification/regression model for male/female of Asian/Caucasian. We evaluated the SCUT-FBP5500 dataset for FBP using different combinations of feature and predictor, and various deep learning methods. The results indicates the improvement of FBP and the potential applications based on the SCUT-FBP5500.
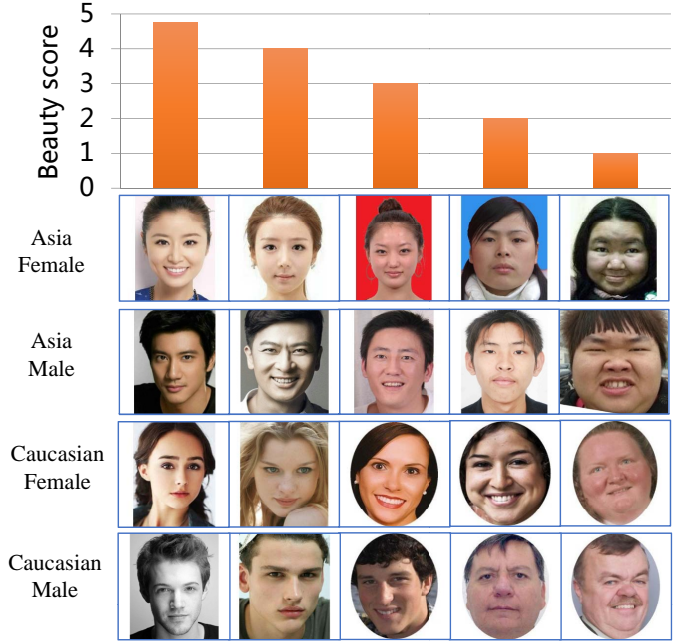
Fig. 1. The images with different facial properties and beauty scores from the proposed SCUT-FBP5500 benchmark dataset. The dataset download URL is shown below the title.

## I. INTRODUCTION

Assessing facial beauty seems natural for human being, but an absolute definition of facial beauty remains elusive. Recently, facial beauty prediction (FBP) have attracted ever-growing interest in the pattern recognition and machining learning communities [8]–[10], which aims to achieve automatic human-consistent facial attractiveness assessment by a computational model. It has application potential in facial makeup synthesis/recommendation [10]–[12], content-based image retrieval [16], aesthetic surgery [10], or face beautification [4]–[7].

From the computational perspective, FBP is still a challenging problem. It is involved with the formulation of visual representation and predictor for the abstract concept of facial beauty. To tackle this problem, various data-driven models, were introduced into FBP. One line of the works follows the classic pattern recognition process, which constructs the FBP system using the combination of the hand-crafted features and the shallow predictors. The related hand-crafted feature derived from visual recognition includes the geometric features, like the geometric ratios and landmark distances [8], [15], [19]–[21], [23], and the texture features, like the Gabor-/SIFT-like features [24]–[26]. Then, a shallow FBP predictoris trained by the extracted feature in a statistical manner.

Another line of works is advanced by the reviving of neural networks, especially the stat-of-the-art deep learning [14]. The hierarchial structure of the deep learning model allows to build an end-to-end FBP system that automatically learns both the representation and the predictor of facial beauty simultaneously from the data. Many works indicate that FBP

TABLE I
REPRESENTATIVE DATABASES FOR FACIAL BEAUTY PREDICTION

| Database | Image Num. | Labelers/Image | Beauty Class | Face Property | Face Landmarks | Publicly Available |
|---|---|---|---|---|---|---|
| Y. Eisenthal et al. [15] | 184 | 28 or 18 | 7 | Caucasian Female | × | × |
| F. Chen et al. [22] | 23412 | unknown | 2 | Asian Male/Female | √ | × |
| H. Gunes et al. [35] | 215 | 46 | 10 | Female | √ | × |
| J. Fan et al. [23] | 432 | 30 | 7 | Generated Female | √ | × |
| M. Redi et al. [34] | 10141 | 78-549 | 10 | Multiple (Sampled from AVA [16]) | × | √ |
| SCUT-FBP [1] | 500 | 70 | 5 | Asian Female | √ | √ |
| **SCUT-FBP5500** | 5500 | 60 | 5 | Asian/Caucasian; Male/Female | √ | √ |

based on deep learning is superior to the shallow predictors with hand-crafted facial feature [1]–[3], [29], [30], [33].

Most of current FBP models are data-driven, which makes benchmark dataset become one of the essential elements for FBP. There have been many works of the benchmark datasets [16], [22]–[24], [30], [33]–[35] involved with FBP, but most of these datasets focus on a specific problem with specific computation constrains, as shown in Table I. Yan et al. [24] regarded FBP as a ranking problem, and proposed a dataset with low-resolution images gathered from social networks. Fan et al. [23] focused on the geometry analysis of FBP and proposed a dataset containing computer-generated faces with different facial proportions. The North-east China database [22], Shanghai database [21], Hot-Or-Not database [32], [33], AVA database [16] and re-sampled face subset of AVA database [34] are large-scale databases involved with FBP, where the Northeast China [22] and Shanghai database [21] are limited for geometric facial beauty analysis without attractiveness ratings; Hot-Or-Not database [32], [33] only focuses on the appearance-based FBP; and the AVA database [16], [34] is originally designed for aesthetic analysis of entire images but not the facial attractiveness. In our previous work, Xie et al. [1] published a SCUT-FBP benchmark dataset, which has led to many FBP models [1]–[3], [31], [37], especially the hierarchial CNN-based FBP models with the state-of-the-art deep learning [1]–[3], [37]. Despite the prevalent usage of the SCUT-FBP, it only contains 500 Asian Female faces, which may limit the performance of the data-demanded model for FBP.

We find that FBP have been formulated the recognition of facial beauty as a specific supervised learning problem of classification [15], [20], [27], [30], regression [13], [19], [23], [29], [38] or ranking [24], [25], [37]. It indicates that FBP is intrinsically a computation problem with multiple paradigms. Previous databases built under specific computation constrains would limit the performance and flexibility of the computational model trained on the dataset, and it is difficult to compare different models derived from the dataset with specific computation paradigm. Therefore, this paper argues that FBP is a multi-paradigm computation problem, and proposes a new diverse benchmark dataset, called SCUT-FBP5500, to achieve multi-paradigm facial beauty prediction.

The SCUT-FBP5500 dataset has totally 5500 frontal faces with diverse properties (male/female, Asian/Caucasian, ages) and diverse labels (face landmark, beauty score, beauty score distribution), which allows different computational model with different FBP paradigms, such as appearance-based/shape-based facial beauty classification/regression/ranking model for male/female with Asian/Caucasian. Furthermore, the diverse faces with beauty scores gathered from 60 different labelers can lead to many interesting research, such as cross-culture facial beauty analysis, personalized FBP [36], or automatic face beautification [4]–[7]. Both shallow prediction model with hand-crafted feature and the state-of-the-art deep learning models were evaluated on the dataset, and the results indicates the improvement of FBP and the potential applications by the SCUT-FBP5500.

The main contributions of this paper can be summarized as following:

1) **Dataset.** We propose a new large-scale SCUT-FBP5500 benchmark dataset that has totally 5500 frontal faces with diverse properties and diverse labels, which allows construction of FBP models with different paradigms.
2) **Benchmark Analysis.** We analyze the samples, score labels, labelers and facial landmarks of the SCUT-FBP5500 statistically, and the visualization of the data illustrates the properties of the SCUT-FBP5500.
3) **Facial Beauty Prediction Evaluation.** Both shallow prediction model with hand-crafted feature and deep learning models are trained on the SCUT-FBP5500 for evaluation, and the results indicates the improvement of FBP based on the proposed dataset with better diversity.

## II. CONSTRUCTION OF SCUT-FBP5500 DATASET

### A. Face Images Collection

The SCUT-FBP5500 Dataset contains 5500 frontal, un-occluded faces aged from 15 to 60 with neutral expression. It can be divided into four subsets with different races and gender, including 2000 Asian females, 2000 Asian males, 750 Caucasian females and 750 Caucasian males. Most of the images of the SCUT-FBP5500 were collected from Internet, where some portions of Asian faces were from the DataTang [39] and some Caucasian faces were from the 10k US Adult database [40].
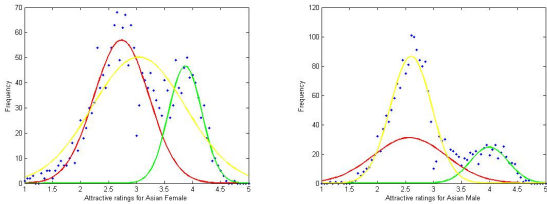
| Subset | CF | CM | AF | AM |
|---|---|---|---|---|
| Total Score Num. | 45000 | 45000 | 120000 | 120000 |
| Outlier Num. | 143 | 181 | 356 | 497 |
| Outlier Portion | 0.3% | 0.4% | 0.3% | 0.4% |
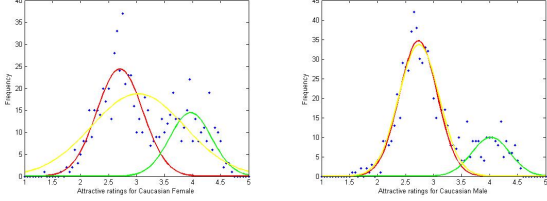


(a) Score distribution of CF (b) Score distribution of CM

(c) Score distribution of AF (d) Score distribution of AM

Fig. 2. Gaussian fitting (yellow curve) and piecewise fitting (red and blue curve) for the visualization of the beauty score distribution of Caucasian female (CF), Caucasian male (CM), Asian female (AF) and Asian male (AM).

## B. Facial Beauty Scores and Facial Landmarks

All the images are labeled with beauty scores ranging from [1, 5] by totally 60 volunteers aged from 18-27 (average 21.6), where the beauty score 5 means most attractive and so on. We developed a web-based GUI system to obtain the facial beauty scores. The labeling system was deployed on the Ali Cloud, and the labeling tasks are distributed to each volunteer in crowdsourcing manners. The four subset, Asian male/female and Caucasian male/female, are labeled separately, where each face of the subset are randomly shown to the volunteer. Then, the volunteer are asked to select the beauty scores within [1, 5] for the face. To reduce the variance in the labeling process, about 10% faces recurred randomly. If the correlation coefficient of the two beauty score of the same faces is less than 0.7, the volunteer would be asked to rate this face once more to decide the final score.

To allow geometric analysis of facial beauty, 86 facial landmarks are located to the significant facial components of each images, such as eyes, eyebrows, nose and mouth. A GUI landmarks location system is developed, where the original location of the landmarks are initialized by the active shape model (ASM) trained by the SCUT-FBP dataset. Then, the detected landmarks by ASM are modified manually by volunteers to ensure the accuracy.
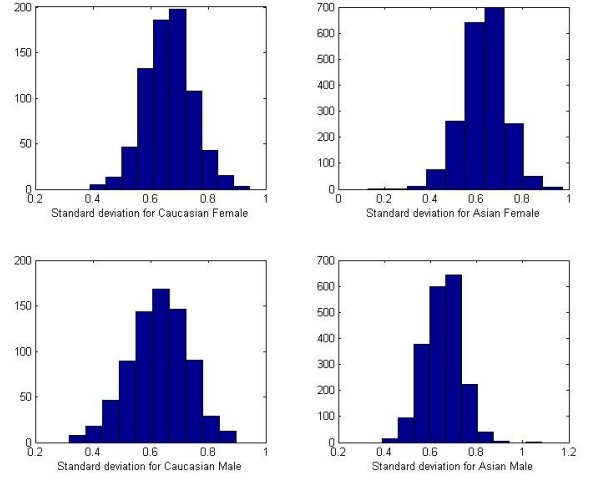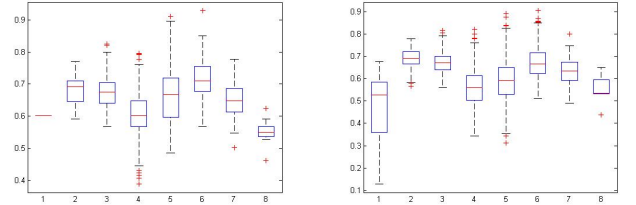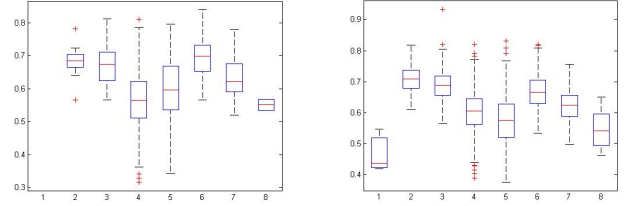


Fig. 3. Distribution of standard deviations of Caucasian female, Asian female, Caucasian male and Asian male, respectively.



(a) Box figure of CF (b) Box figure of AF

(c) Box figure of CM (d) Box figure of AM

Fig. 4. Box figures of standard deviations of Caucasian female (CF), Asian female (AF), Caucasian male (CM) and Asian male (AM), respectively.

## III. BENCHMARK ANALYSIS OF THE SCUT-FBP5500

We made benchmark analysis of the beauty scores, labelers and face landmarks of the SCUT-FBP5500 with different gender and races, including Asian female (AF), Asian male (AM), Caucasian female (CF) and Caucasian male (CM).

## A. Distribution of Beauty Scores

We visualize the distribution of the beauty scores of AF, AM, CF and CM, respectively. To obtain better visualization, we preprocess the data and filter the outliers of the beauty scores. We regard the average score of all the 60 labelers as the ground-truth. If the score of specific labeler for the same face differs from the ground-truth over 2, the score is treated as outlier and is removed from the distribution visualization.

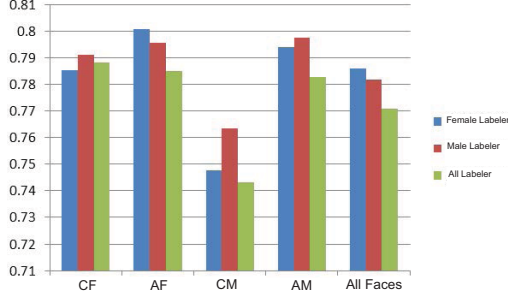|                | CF    | AF    | CM    | AM    | All Faces |
|----------------|-------|-------|-------|-------|-----------|
| Female Labelers | 0.785 | 0.800 | 0.747 | 0.793 | 0.785     |
| Male Labelers   | 0.791 | 0.795 | 0.763 | 0.797 | 0.781     |
| All Labelers    | 0.788 | 0.785 | 0.743 | 0.782 | 0.770     |



Fig. 5. Correlation Coefficient of male and female labelers for beauty score of CF, AF, CM, AM and all the faces in SCUT-FBP5500.



(a) PCA Analysis of AF



(b) PCA Analysis of AM

Fig. 6. PCA analysis of face landmarks of Asian female (AF) and Asian male (AM).

The number and portion of the outliers in the four subset are listed in Table II, and the small portion of outlier indicate the reliability of the labeling process for beauty score.

Since the outlier portion of the beauty score is tiny, the distributions of the original data and the preprocessed data is mostly similar. Therefore, we visualized the score distribution of the SCUT-FBP5500 using the preprocessed data for all the four subset, respectively. Two distribution fitting schemes are used: one is Gaussian fitting (yellow curve), the other is piecewise fitting (red and blue curve), as shown in Fig. 2. The results indicates that the beauty scores of all the four subset can be approximately fitted by a mixed distribution model with two Gaussian components.

### B. Standard Deviation of Beauty Scores

We calculate the standard deviation of the scores gathered from different labelers to the ground-truth, and illustrate the results as histogram in Fig. 3 and as box figure in Fig. 4. We observe that the distribution of standard deviations is similar to Gaussian distribution, and most of the standard deviations are within a reasonable range of [0.6, 0.7].

### C. Correlation of Male/Female Labelers

In this subsection, we investigate the correlation between the male and female Asian labelers for the beauty scores, as shown in Table III and Fig. 5. We observe that the correlation of Asian faces is persistently larger than Caucasian. It is consistent to the psychological research that human have better facial beauty perception for the faces from the same race.

### D. PCA Analysis of Facial Geometry

We visualize the 86-points face landmarks of the dataset using principle component analysis (PCA). Fig. 6 illustrate the mean and the five first principle component of the facial geometry of Asian female (AF) and Asian male (AM), where the landmarks data of Caucasian share similar distribution to Asian faces. We observe that the face shape is one of the main component influence the face geometry of beauty, which is consistent to related psychological research and previous works in [21], [22]

## IV. FBP EVALUATION VIA HAND-CRAFTED FEATURE AND SHALLOW PREDICTOR

This section, we evaluate the SCUT-FBP5500 using the hand-crafted feature with shallow predictor, while the next section introduce some state-of-the-art deep learning model to achieve FBP.

### A. Geometric Feature with Shallow Predictor

We extract a 18-dimensional ratio feature vector from the faces and formulate FBP based on different regression models, such as the linear regression (LR), Gaussian regression (GR), and support vector regression (SVR). Comparison were performed for Caucasian female/male and Asian female/male subsets, and the performance of different model are measured using pearson correlation coefficient (PC) [15], maximum absolute error (MAE) and root mean square error (RMSE) after 10 folds cross validation. The results are listed in Table IV and Table V, which can be regarded as a baseline for the geometric analysis of FBP.

| | Asian Female | | | Asian Male | | |
|---|---|---|---|---|---|---|
| | LR | GR | SVR | LR | GR | SVR |
| PC | 0.6771 | 0.7057 | 0.7008 | 0.6348 | 0.6923 | 0.6816 |
| MAE | 0.402 | 0.387 | 0.3876 | 0.3894 | 0.3572 | 0.356 |
| RMSE | 0.5246 | 0.5057 | 0.5089 | 0.5085 | 0.4752 | 0.4823 |

| | Caucasian Female | | | Caucasian Male | | |
|---|---|---|---|---|---|---|
| | LR | GR | SVR | LR | GR | SVR |
| PC | 0.6809 | 0.7263 | 0.7093 | 0.6063 | 0.63 | 0.6397 |
| MAE | 0.3986 | 0.3862 | 0.4001 | 0.3871 | 0.3689 | 0.3617 |
| RMSE | 0.5239 | 0.4908 | 0.5087 | 0.4899 | 0.4784 | 0.4739 |

#### TABLE V
FACIAL BEAUTY PREDICTION USING GEOMETRIC FEATURE WITH SHALLOW MODELS FOR THE WHOLE DATASET

| | Linear Regression | Gaussian Regression | SVR |
|---|---|---|---|
| PC | 0.5948 | 0.6738 | 0.6668 |
| MAE | 0.4289 | 0.3914 | 0.3898 |
| RMSE | 0.5531 | 0.5085 | 0.5132 |

#### TABLE VI
FACIAL BEAUTY PREDICTION USING GABOR FEATURE WITH TWO SAMPLING SCHEME ON WHOLE DATASET

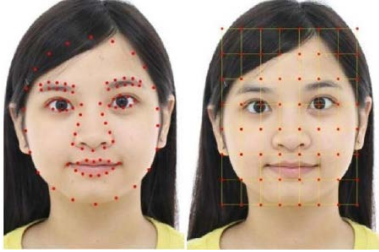| | 86-keypoints | | 64UniSample | |
|---|---|---|---|---|
| | GR | SVR | GR | SVR |
| PC | 0.7472 | 0.6691 | 0.6764 | 0.8065 |
| MAE | 0.3554 | 0.3891 | 0.4014 | 0.3976 |
| RMSE | 0.4599 | 0.5065 | 0.5177 | 0.5126 |



Fig. 7. Two sampling schemes to extract the appearance feature of FBP, where the left one is the 86-keypoints method and the right one is the UniSamplePoint method.

### B. Appearance Feature with Shallow Predictor

We extract 40 Gabor feature maps from every original image in five directions and eight angles. Then, we obtain the appearance feature of FBP using two different sampling schemes that extracts some component of the Gabor feature maps as following:

- Sample 86-keypoints from each 40 Gabor feature maps to obtain a 3340-dimensional feature vector, as shown in the right sub-figure of Fig. 7;
- Use 64UniSample to obtain a 2560-dimensional feature vector, as shown in the left sub-figure of Fig. 7.

Finally, we use PCA to reduce the extracted feature dimension before we train the predictor. The results of the appearance-based shallow predictors for all the data are shown in Table VI.

#### TABLE VII
COMPARISON OF ALEXNET [41], RESNET-18 [42] AND RESNEXT-50 [43] IN MEASUREMENT OF PC, MAE AND RMSE BY 5-FOLD CROSS VALIDATION

| PC | 1 | 2 | 3 | 4 | 5 | Average |
|---|---|---|---|---|---|---|
| AlexNet | 0.8667 | 0.8645 | 0.8615 | 0.8678 | 0.8566 | 0.8634 |
| ResNet-18 | 0.8847 | 0.8792 | 0.8929 | 0.8932 | 0.9004 | 0.89 |
| ResNeXt-50 | **0.8985** | **0.8932** | **0.9016** | **0.899** | **0.9064** | **0.8997** |
| **MAE** | 1 | 2 | 3 | 4 | 5 | Average |
| AlexNet | 0.2633 | 0.2605 | 0.2681 | 0.2609 | 0.2728 | 0.2651 |
| ResNet-18 | 0.248 | 0.2459 | 0.243 | 0.2383 | 0.2383 | 0.2419 |
| ResNeXt-50 | **0.2306** | **0.2285** | **0.226** | **0.2349** | **0.2258** | **0.2291** |
| **RMSE** | 1 | 2 | 3 | 4 | 5 | Average |
| AlexNet | 0.3408 | 0.3449 | 0.3538 | 0.3438 | 0.3576 | 0.3481 |
| ResNet-18 | 0.3258 | 0.3286 | 0.3184 | 0.3107 | 0.2994 | 0.3166 |
| ResNeXt-50 | **0.3025** | **0.3084** | **0.3016** | **0.3044** | **0.2918** | **0.3017** |

#### TABLE VIII
COMPARISON OF ALEXNET [41], RESNET-18 [42] AND RESNEXT-50 [43] IN MEASUREMENT OF PC, MAE AND RMSE BY 60% TRAINING AND 40% TESTING

| | AlexNet | ResNet-18 | ResNeXt-50 |
|---|---|---|---|
| PC | 0.8298 | 0.8513 | **0.8777** |
| MAE | 0.2938 | 0.2818 | **0.2518** |
| RMSE | 0.3819 | 0.3703 | **0.3325** |

## V. FBP EVALUATION VIA DEEP PREDICTOR

We evaluate three recently proposed CNN models with different structures for FBP, including AlexNet [41], ResNet-18 [42] and ResNeXt-50 [43]. All these CNN models are trained by initializing weights using networks pre-trained on the ImageNet dataset. The evaluation of were performed under two different experiment settings as following:

1) The models were trained and tested using 5-fold cross validation, which means each fold containing 20% samples (1100 images). The accuracy of each fold and the average of all the 5 fold are shown in Table VII.
2) The models were trained using 60% samples (3300 images), and tested with the rest 40% (2200 images). The results are shown in Table VIII.

The results illustrates that the deepest CNN-based ResNeXt-50 model [43] obtains the best performance comparing to the ResNet-18 and AlexNet in both the experiment setting. It can be observed that all the deep CNN model are superior to the shallow predictor with hand-crafted geometric feature in Table V or appearance feature in Table VI. It indicates the effectiveness and powerfulness of the end-to-end feature learning deep model for FBP.

Comparing the results of Table VII and Table VIII, we also find that the accuracy of all the 5-fold cross validation is slightly higher than the results of the split of 60% training and 40% testing. One of the reasons may be due to the amounts and diversity of the training samples, since the 5-fold cross validation use 80% samples to train the models. This observation indicates that the data augmentation techniques may further improve the performance of the deep FBP model, which merits exploring in the future.

## VI. Conclusion

In this paper, we introduce a new diverse benchmark dataset, called SCUT-FBP5500, to achieve multi-paradigm facial beauty prediction. The SCUT-FBP5500 dataset has totally 5500 frontal faces with diverse properties (male/female, Asian/Caucasian, ages) and diverse labels (face landmarks, beauty scores within [1, 5], beauty score distribution). Benchmark analysis have been made for the beauty scores and landmarks in SCUT-FBP5500, and the visualization of the data shows the statistical properties of the dataset. Since the SCUT-FBP5500 is designed for multi-paradigm, it can be adapted to different FBP models for different tasks, like appearance-based or shape-based model for facial beauty classification/regression/ranking. We evaluated the SCUT-FBP5500 using different combinations of feature and predictor, and deep learning models, where the results indicates the reliability of the dataset.

## References

[1] D. Xie, L. Liang, L. Jin, J. Xu and M. Li, "SCUT-FBP: A Benchmark Dataset for Facial Beauty Perception," in *Proc. of IEEE SMC*, pp. 1821-1826, 2015.

[2] L. Liang, D. Xie, L. Jin, J. Xu, M. Li, and L. Lin, "Region-aware scattering convolution networks for facial beauty prediction," in *Proc. of IEEE ICIP*, pp. 2861-2865, 2017.

[3] J. Xu, L. Jin, L. Liang, Z. Feng, D. Xie, and H. Mao, "Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN)," in *Proc. of IEEE ICASSP*, pp. 1657-1661, 2017.

[4] L. Liang, L. Jin, and D. Liu, "Edge-aware label propagation for mobile facial enhancement on the cloud," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 27, no. 1, pp. 125-138, 2017.

[5] L. Liang, L. Jin, and X. Li, "Facial skin beautification using adaptive region-aware mask," *IEEE Trans. on Cybernetics*, vol. 44, no. 12, pp. 2600-2612, 2014.

[6] J. Li, X. Chao, L. Liu, X. Shu, and S. Yan, "Deep face beautification," *Proc. ACM International Conference on Multimedia*, pp. 793-794, 2015.

[7] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "Data-driven enhancement of facial attractiveness," *ACM Trans. Graph.*, pp. 28:1-10, 2008.

[8] H. Gunes, "A survey of perception and computation of human beauty," in *Proc. of J-HGBU*, pp. 19-24, 2011.

[9] A. Laurentini and A. Bottino, "Computer analysis of face beauty: A survey," *Comput. Vision and Image Underst.*, vol. 125, pp. 184–199, 2014.

[10] D. Zhang, F. Chen, and Y. Xu, *Computer models for facial beauty analysis*. Springer International Publishing Switzerland, 2016.

[11] L. Liu, J. Xing, S. Liu, H. Xu, X. Zhou, and S. Yan, "Wow! you are so beautiful today!," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 11, no. 1s, p. 20, 2014.

[12] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz and H. Seidel, "Computer-suggested facial makeup," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 485-492, 2011.

[13] Y. Mu, "Computational facial attractiveness prediction by aesthetics-aware features," *Neurocomputing*, vol. 99, pp. 59-64, 2013.

[14] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 251, pp. 436-444, 2015.

[15] Y. Eisenthal, G. Dror and E. Ruppin, "Facial Attractiveness: Beauty and the Machine," *Neural Computation*, vol. 18, pp. 119-142, 2006.

[16] N. Murray, L. Marchesotti L, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. of CVPR*, pp. 2408-2415, 2012.

[17] A. Laurentini and A. Bottino, "Computer analysis of face beauty: A survey," *Computer Vision and Image Understanding*, vol. 125, pp. 184-199, 2014.

[18] I. Stephen, M. Law Smith, M. Stirrat, and D. Perrett, "Facial skin coloration affects perceived health of human faces," *International Journal of Primatology*, vol. 30, no. 6, pp. 845-857, 2009.

[19] A. Kagian, G. Dror, T. Leyvand, D. Cohen-Or, and E. Ruppin, "A humanlike predictor of facial attractiveness," in *Proc. of NIPS*, pp. 649-656, 2006.

[20] H. Mao, L. Jin and M. Du, "Automatic classification of Chinese female facial beauty using Support Vector Machine," in *Proc. of IEEE SMC*, pp. 4842-4846, 2009.

[21] D. Zhang, Q. Zhao and F. Chen, "Quantitative analysis of human facial beauty using geometric features," *Pattern Recognition*, vol. 44, no. 4, pp. 940-950, 2011.

[22] F. Chen and D. Zhang, "A benchmark for geometric facial beauty study," *Medical Biometrics*, pp. 21-32, 2010.

[23] J. Fan, K. P. Chau, X. Wan, L. Zhai and E. Lau, "Prediction of facial attractiveness from facial proportions," *Pattern Recognition*, vol. 45, pp. 2326-2334, 2012.

[24] H. Yan, "Cost-sensitive ordinal regression for fully automatic facial beauty assessment," *Neurocomputing*, no. 129, pp. 334-342, 2014.

[25] H. Altwaijry and S. Belongie, "Relative ranking of facial attractiveness," in *IEEE Workshop on WACV*, pp. 117-124, 2013.

[26] Y. Chen, H. Mao, L. Jin, "A novel method for evaluating facial attractiveness," in *IEEE Proc. ICALIP*, pp. 1382-1386, 2010.

[27] W. Chiang, H. Lin, C. Huang, L. Lo, and S. Wan, "The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D Moir features," *Pattern Recognition*, vol. 47, no. 3, pp. 1249-1260, 2014.

[28] S. Kalayci, H. K. Ekenel, and H. Gunes, "Automatic analysis of facial attractiveness from video," in *Proc. of ICIP*, pp. 4191-4195, 2014.

[29] J. Gan, L. Li, Y. Zhai and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, no. 144, pp. 295-303, 2014.

[30] S. Wang, M. Shao and Y. Fu, "Attractive or not? Beauty prediction with attractiveness-aware encoders and robust late fusion," *ACM Multimedia*, pp. 805-808, 2014.

[31] Y. Ren and X. Geng, "Sense beauty by label distribution learning," in *Proc, IJCAI*, pp. 2648–2654, 2017.

[32] R. White, A. Eden, and M. Maire, "Automatic prediction of human attractiveness," *UC Berkeley CS280A Proj.*, vol. 1, p. 2, 2004.

[33] D. Gray, K. Yu, W. Xu and Y. Gong, "Predicting facial beauty without landmarks," in *Proc. ECCV*, 2010.

[34] M. Redi, N. Rasiwasia, G. Aggarwal and A. Jaimes, "The Beauty of Capturing Faces: Rating the Quality of Digital Portraits," in *IEEE Workshops on AFGR*, pp. 1-8, 2015.

[35] H. Gunes and M. Piccardi, "Assessing facial beauty through proportion analysis by image processing and supervised learning," *International Journal of Human-Computer Studies*, vol. 64, no. 12, pp. 1184-1199, 2006.

[36] J. Whitehill and J.R. Movellan, "Personalized facial attractiveness prediction," in *Proc. AFGR*, 2008.

[37] Y. Fan, S. Liu, B. Li, Y. Fan, Z. Guo, A. Samal, J. Wan and Stan Z. Li "Label Distribution based facial attractiveness computation by deep residual learning", *IEEE Transactions on Multimedia*, 2017 (Online).

[38] B. Davis and S. Lazebnik, "Analysis of human attractiveness using manifold kernel regression," in *Proc. ICIP*, 2008.

[39] "DataTang," url: http://datatang.com/.

[40] W.A. Bainbridge, P. Isola and A. Oliva, "The intrinsic memorability of face images." *Journal of Experimental Psychology: General*, vol. 142, no. 4, pp. 1323-1334, 2013.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc, NIPS*, 2012, pp. 1097–1105.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc, CVPR*, 2016, pp. 770–778.

[43] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *arXiv preprint arXiv:1611.05431*, 2016.