

Aesthetic Quality Assessment of Photos with Faces

Weining Wang¹, Jiexiong Huang¹, Xiangmin Xu^{1(✉)}, Quanzeng You², and Jiebo Luo²

¹ South China University of Technology, Guangzhou, China
{wnwang, xmxu}@scut.edu.cn, h.jiexiong@mail.scut.edu.cn

² University of Rochester, Rochester, NY 14627, USA
{qyou, jluo}@cs.rochester.edu

Abstract. Aesthetic quality assessment of photos is a challenging task regarding the complexity of various photos and subjectivity of human's aesthetic perception. Recent research has suggested that photos with different contents have different aesthetic characters. However, these different aesthetic characters are not considered in previous work of aesthetic assessment. Meanwhile, photos with human faces have become increasingly popular and constitute an important part of social photo collections. In this work, we analyze the characters of this particular category of photos, human faces, and study the impact of them on aesthetic quality estimation. This study could have many potential applications, including selection of high aesthetic face photos, face photo editing and so on. To solve this problem, we design new handcrafted features and fine-tuned a new deep Convolutional Neural Network (CNN) for features. Next, we build decision fusion model to employ all the proposed features for aesthetic estimation. In addition, we analyze the effectiveness of different groups of features in a face photo classification task to better understand their differences. Experimental results show that our proposed features are effective and the classifier outperforms several up-to-date approaches in aesthetic quality assessment.

Keywords: Aesthetic quality assessment · Faces
Deep Convolutional Neural Network · Decision fusion

1 Introduction

Aesthetic quality assessment is a subjective task, which is related to human's perception of visual stimuli. It can improve user experiences and provide better quality of service in many applications, including image retrieval, photo editing, photography and so on. Automated aesthetic quality assessment tries to automatically predict human's aesthetic perception. It is a challenging task regarding the complexity of various photos and subjectivity of human's aesthetic perception. Recent work on automated aesthetic quality assessment has been focusing on designing robust machine learning algorithms to distinguish the aesthetic quality of photos.

Aesthetic features are critical and directly impact the performance of current machine learning models on this task. Researchers have spent huge efforts in designing novel and descriptive aesthetic features, inspired from painting, photography, art, and so on [1–3, 9–11, 17]. Meanwhile, they also try to provide additional solutions including generic

features [4] and geo-content information [5]. More recently, deep learning method has shown significantly improved results on a wide range of computer vision tasks [6–8, 15]. It has also been applied to automated aesthetic assessment. Lu et al. [8] designed a double-column Convolutional Neural Networks (CNNs) for aesthetic quality categorization. Dong et al. [7] used a trained deep learning network to extract features of photos and then applied the Support Vector Machine (SVM) to estimate the aesthetic quality of photos. However, these studies have been using the CNNs fine-tuned from general image classification tasks and studying the aesthetic quality of general images. Meanwhile, Tang’s research [10] showed that photos with different contents have different aesthetic characters. More than that, psychology research in perception also confirms that certain kinds of contents will be more attractive than others to our eyes, either because we have learned to expect more information from them or because they appeal to our emotions or desires [11]. Therefore, it is expected that by designing different aesthetic features for different kinds of photos, we can achieve better performance in accessing photo aesthetic quality.

In this paper, our research focuses on photos with human faces. They constitute an important part of social photo collections. Among our 500 randomly selected photos from online social network, the percentage of photos with faces is 45.6%, which is also the largest category. Indeed, this coincides with the finding that pictures with faces can attract more likes, comments and attention from online social network users¹. Therefore, the study of aesthetic quality on this particular category of photos can still have many potential applications.

Indeed, there have been several previous studies focusing on aesthetics of photos with faces. Li et al. [11] built a dataset of photos with faces, where each photo has been assigned aesthetic scores by multiple annotators. They also evaluated the performance of several different categories of aesthetic features, including pose features, location features and composition features. However, the dataset is too small to validate the effectiveness and robustness of different groups features. Tang et al. [10] built another dataset with 7 different categories, where “human” is one of them. Each photo has been assigned a label of “high quality” or “low quality”. Particularly, they incorporated face specific features for photos with faces. However, to the best of our knowledge, no research is conducted on employing deep learning to study aesthetic quality of photos with faces. One of the main difficulties is the lack of a large scale labeled datasets.

To solve the above mentioned problems, we make the following contributions:

1. Based on large-scale AVA dataset [12], we collect a dataset of photos with faces AVA_Face. It consists of 20,320 photos, which is large enough to fully evaluate different features.
2. We design 78 facial aesthetic features, including facial composition features, Eyes-Chambers features, shadow features, expression features and saliency features.
3. We apply CNN to learn aesthetic features due to the availability of a relative large scale dataset.

¹ <http://www.news.gatech.edu/2014/03/20/face-it-instagram-pictures-faces-are-more-popular>.

- We fuse the features from CNN and the other two kinds of handcrafted aesthetic features using decision fusion. This is different from previous researches, where the main strategy is early fusion.

The overall framework of our method is shown in Fig. 1. The first component is feature extraction. It includes both the hand crafted features as well as features from a fine-tuned CNN. Next, we use decision fusion to aggregate the prediction result from the different features and to evaluate the aesthetic quality.

The paper is organized as follows. The newly designed facial aesthetic features are introduced in Sect. 2. In Sect. 3, we apply a new CNN model to learn aesthetic features. In Sect. 4, we introduce Decision Fusion. The datasets and experiments are presented in Sect. 5, then we conclude the paper in Sect. 6.

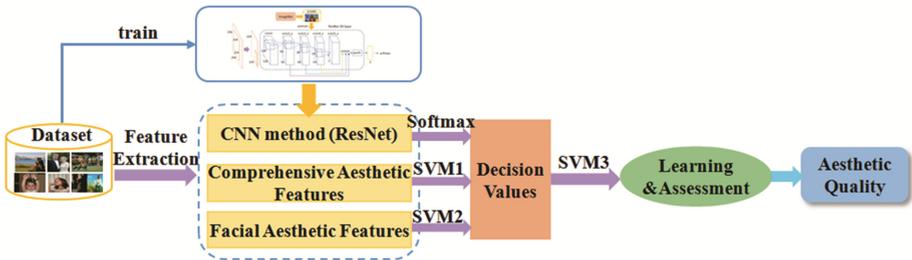


Fig. 1. Overview of the proposed work.

2 Facial Aesthetic Features

In photos with faces, people pay much more attention to the face regions. Intuitively, features related to face are closely related the aesthetics of photos. Therefore, we design 78 dimensions of facial aesthetic features. (1) For multiple face regions in the photo, 47 facial composition features and 5 facial saliency features are extracted. (2) For the largest face region, 12 Eyes-Chambers features, 9 facial shadow features and 5 facial expression features are extracted.



Fig. 2. Examples of the detected key points of faces.

Face++ Research Toolkit v0.1 [19] is used to detect face and their key points in this paper. Figure 2 shows some examples of the detected face key points, which are going to help us design and extract facial aesthetic features.

2.1 Facial Composition Features

We define two kinds of facial composition features as follows.

Facial distribution features: the number of faces f_1 , the area of the largest face f_2 , the average area of all faces f_3 , the standard derivation of areas of faces f_4 , the proportion of the area of faces f_5 , the maximum distance between faces f_6 , the minimum distance between faces f_7 , the average distance of faces f_8 , and the compactness of face f_9 . Minimal Spanning Tree algorithm [16] is used to calculate the distances between faces. In calculating compactness of face, we first draw a smallest rectangle region, which can contain all faces, and then calculate the diagonal length of this rectangle. Next, we obtain the compactness of face by dividing the average distance to the computed diagonal length.

Facial basic features: The facial features consist of color and texture consistency features. In all face regions, the mean value and standard derivation of hue $f_{10} \sim f_{11}$, saturation $f_{12} \sim f_{13}$, and value $f_{14} \sim f_{15}$ are extracted as the color consistency features. Then, we use Gabor wavelet with 4 scales $\sigma \in \{0, 3, 6, 9\}$ and 4 directions $\theta \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ\}$ to extract facial texture features. For each pair of scale and direction, we compute their averages as $f_{16} \sim f_{31}$ and standard derivations as $f_{32} \sim f_{47}$. Both of the averages and the derivations are known as the texture consistency features.

2.2 Facial Saliency Features

Salient regions are important in aesthetics of photos [2, 10]. In photos with faces, faces, which are the main saliency regions, attract the most attention. Therefore, we detect face regions and extract facial clarity and complexity as facial saliency features.

We calculate facial clarity features f_{48} as follows [10]:

$$C_{face} = \{(x, y) \mid |F_{face}(x, y)| > \beta \max(F_{face}(x, y))\} / Num \quad (1)$$

where Num is the number of pixels belong to the face region, F_{face} is the Fourier transform of face regions, and β is the threshold (set to 0.2 in this paper).

Then, we define f_{49} as the ratio of the clarity of face regions to the entire photo.

$$R_{face} = C_{face} / C_{photo} \quad (2)$$

To extract facial complexity features in face regions, we use the number of pixels to describe it. We segment the super pixels of a total photo and get the number of super pixels of face regions N_{face} and the number of super pixels of the background N_{unface} .

We define I_{face} and I_{unface} are the sets of pixels of face regions and the other regions, then we calculate the facial complexity features $f_{50} \sim f_{52}$ as follows [10].

$$Complexity_1 = N_{face} / \|I_{face}\| \quad (3)$$

$$Complexity_2 = N_{unface} / \|I_{unface}\| \quad (4)$$

$$Complexity_3 = N_{face} / \|I_{unface}\| \quad (5)$$

2.3 Eyes-Chambers Features

Researchers have tried to design aesthetic features inspired by the locations of different facial parts [13]. Comparably, we design Eyes-Chambers features considered the aesthetic influence.

Five Eyes and Three Chambers are traditional criteria for beautiful faces in Chinese Physiognomy. Five Eyes requires that the width of our two eyes, the width between two eyes and distances between two eyes and the contours of face are all about 1/5 of the face width. Three Chambers requires that the length of nose, the distance between nose and jaw and the distance between nose and hair line are all about 1/3 of the face length. It revealed that the width and distance between organs are important determinants of beautiful faces.

We calculate the width of two eyes $f_{53} \sim f_{54}$, the portion of them to the width of face $f_{55} \sim f_{56}$, the distance between two eyes and the portion of it to the width of face $f_{57} \sim f_{58}$, and the ratio of the width of two eyes f_{59} as the Eyes features. Similarly, we calculate length of nose f_{60} , the distance between nose to jaw f_{61} , the ratio of them f_{62} and the portion of them to the height of face $f_{63} \sim f_{64}$ as the Chambers features.

2.4 Facial Shadow Features

The study from [14] indicates that the contrasts between brightness and darkness are able to highlight major regions of photos and thus is correlated with aesthetic qualities. Figure 3 shows several examples, where shadows are correlated with the quality of photos. In general, we can estimate the bright regions and dark regions of faces and use them as shadow computational templates to qualify the shadow values as the facial shadow features. The main steps are shown in Fig. 4.

As shown in Fig. 4(c), the two sub regions with the same color are one pair of shadow computational templates. In total, we have 9 pairs of shadow computational templates. For each pair, we calculate the ratio of brightness of two sub region as the shadow features. In such a way, we receive 9 facial shadow features $f_{65} \sim f_{73}$. Equation (6) shows how to compute the shadow feature for a given pair of sub regions.

$$TL_k = \frac{\sum_{i \in T_{k1}} v(i)}{\left| \left\{ v(i) \mid i \in T_{k1} \right\} \right|} / \frac{\sum_{i \in T_{k2}} v(i)}{\left| \left\{ v(i) \mid i \in T_{k2} \right\} \right|} \quad k \in \{1, 2, \dots, 9\} \quad (6)$$

where T_{k1} and T_{k2} are the two sub regions in the k th pair of shadow computational template, i is the pixel of photos, $v(i)$ is the value of the pixel i in HSV.



Fig. 3. Difference of shadow in photos with faces.

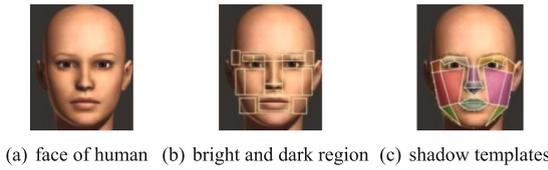


Fig. 4. Shadow computational templates.

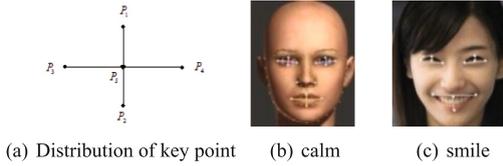


Fig. 5. Examples of facial expression features.

2.5 Facial Expression Features

Facial expressions of a given photo can affect viewer’s subjective evaluation of its aesthetic quality [23]. In general, active and positive facial expressions are expected to receive higher aesthetic responses from the viewers. As shown in Fig. 5, facial expressions are closely related to the detected facial key points. In particular, EyeShape₁, EyeShape₂, MouthShape₁, MouthShape₂ are calculated as the facial expression features $f_{74} \sim f_{77}$. The mean of EyeShape₁ and EyeShape₂ is also included as f_{78} . See Eqs. (7), (8), (9) and (10) for detailed computation of these features.

$$EyeShape_1 = \frac{ED(P_1, P_2)}{ED(P_3, P_4)} \quad (7)$$

$$EyeShape_2 = \frac{ED(P_1, P_5)}{ED(P_5, P_2)} \quad (8)$$

$$MouthShape_1 = \frac{ED(P'_1, P'_2)}{ED(P'_3, P'_5)} \quad (9)$$

$$MouthShape_2 = \frac{ED(P'_1, P'_5)}{ED(P'_5, P'_2)} \quad (10)$$

where $ED(...)$ is the Euclidean distance, P stands for the set of key points around eyes and P' stands for the set of key points around mouth (see Fig. 5 (a) for more details).

3 Aesthetic Quality Assessment Using Deep Learning Method

Recently, deep learning method achieved significant performance improvement in many computer vision tasks [6–8, 15, 22]. From its success in ImageNet Challenge [15, 18], Convolutional Neural Network has been widely used to solve other challenging computer vision tasks. In particular, pre-trained CNN models on ImageNet have been broadly employed for image classification, which has shown promising results on many related vision problems.

In this work, we fine-tune a new CNN model on the pre-trained ResNet, which performed best on the ILSVRC2015 classification task [15]. We choose ResNet-50-layer model as the basic structure and replace the number of neurons in fully connected layer as 2, which is the number of the aesthetic quality categories. As we know, aesthetic quality assessment of photos need us to combine both the global and local information of photos. Therefore, in our new CNN model, we innovatively pool the feature maps of the second, third and fourth convolution layer ($conv2_x$, $conv3_x$, $conv4_x$) to combine with the feature maps of fifth convolution group layer ($conv5_x$) for learning. Figure 6 shows the proposed deep CNN architecture for aesthetic quality assessment.

As shown in Fig. 6, though photos will be resized into the same size when they are input into CNN, it may cause a photo loss its composition information if its width is different to its height. To solve it, we pad each photo to make its width is equal to its height.

In this paper, we fine-tune our model by the pre-trained ResNet-50-layer model on Caffe [24]. When fine-tuning the network, the learning rate of the convolution layers is 0.001 and for the fully connected layer, the learning rate is 0.005. The learning rates will decrease 90% after 7 epochs. The outputs of softmax will be considered as features learned by CNN in decision fusion which will be introduced in Sect. 4. The deep CNN performs very well in aesthetic assessment which will be discussed in detail in Sect. 5.

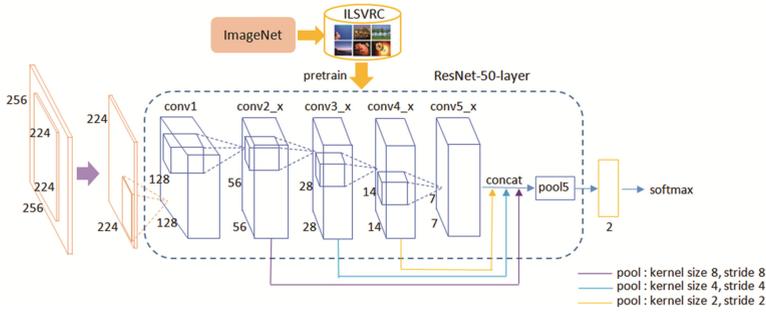


Fig. 6. The construction of CNN in this paper.

4 Decision Fusion

Comprehensive features have been employed for evaluation of photo aesthetic quality [2]. There are 86 comprehensive aesthetic features including 56 low-level features, 11 rule-based features, 6 information theory features and 13 visual attention features [2]. The results from [2] suggest the effectiveness of these features in aesthetic evaluation of photos. Therefore, we extract these features in our experiments as well.

As we introduce above, CNN performs well in image classification. However, it is usually applied as a black box and ignore the basic information of photos to for computer vision tasks. After extracting facial aesthetic features and comprehensive features, how to combine the advantages of handcrafted features and CNN method to enhance the performance of aesthetic assessment is worth considering.

In this paper, we use decision fusion method [20] to fuse them (see Fig. 1). First, we learn different SVM classifiers using two different groups of aesthetic handcrafted features. Each SVM is a binary classifier, which classify the photos into high or low aesthetic quality group. For CNN architecture, a softmax classifier is used to classify photos into high or low aesthetic quality as well. Then, the decision values of each photo from the three classifiers, which are considered as features, are used inputs for decision fusion. Specifically, decision fusion learns another binary SVM classifier to produce the final result. The results show that decision fusion greatly improves the performance of aesthetic assessment.

5 Datasets and Experiments

5.1 Datasets

There are two widely used datasets of photos with faces including the Li’s human dataset [11] and the human category of CUHKPQ [10]. Li’s human dataset [11] consists of 500 photos along with their aesthetic scores. CUHKPQ [10] is a database consists of 17,690 photos with less noise, which has 7 different categories. In “human” category, there are 3,148 photos with a label of high quality or low quality.

AVA [12] is a large scale database, which consists of 255,529 images with aesthetic scores. Based on AVA dataset, we collect a new dataset AVA_Face², which consists of photos with faces. This dataset has 20,320 photos, where each photo has at least one human face. Following Dong’s [7] method, we use the following two ways to obtain the binary labels for each photo:

AVA_Face1: The photos with scores higher than 5 will get labels of “good”, and others will get labels of “bad”. Eventually, there are 15,017 photos in “good” category and 5,303 photos in “bad” category.

AVA_Face2: The photos rating in top 10% are in “good” category and the down 10% are in “bad” category. Therefore, there are both 2,032 photos in two categories.

5.2 Experiments

5.2.1 Experiments on the Li’s Dataset

In [11], Li et al. test their features by aesthetic classification and aesthetic regression. In classification, they calculate the accuracy within one Cross-Category Error (CCE) while they calculate the residual sum-of-squares error (RES) in regression.

In our experiments, we extract our proposed aesthetic features on Li’s dataset. Since the dataset is too small to fine-tune CNN model, we use the best model trained on AVA_Face1 to test the photos for the decision values. Then, we fuse these features using decision fusion and then train and test the photos as the way of Li’s experiment on Li’s experiment [11] in both classification and regression. Last, we calculate the accuracy within one CCE and the RES, which shown in Table 1. In Table 1, we can see that in classification, our features perform better and increase the accuracy within one CCE a lot. These two results show that our proposed features perform well in aesthetic evaluation of photos with faces.

Table 1. The results on Li’s dataset.

	Li’s method [11]	Our method
Accuracy within one CCE	68%	93.2%
RES	2.38	2.27

5.2.2 Experiments on Human Category of CUHKPQ Dataset

In paper [10], Tang et al. randomly choose half photos of CUHKPQ to be training set and the others to be testing set, and then repeat this random partition 10 times. On human category of CUHKPQ, we extract all the three groups of features and test the features as the way of Tang’s method [10]. When using CNN method, we randomly choose half photos as training set to train the model and test the others to get their decision values. Then we swap the training set and testing set to get the decision values of all photos.

We compare with two recent studies on this dataset. The first study is from Tang et al. [10] proposed facial aesthetic features and global features in assessment.

² The IDs of the photos in AVA_Face can be downloaded at https://github.com/huangjx07/avaface_id.

The second one is Guo et al. [21], where semantic LLC features are included. The results are in Table 2.

In comparison, we can see that our proposed aesthetic features all perform better than Guo’s features and Tang’s features. The accuracy of our method is higher than another two proposed methods, and in Fig. 7(a), the ROC (receiver operating characteristic) curve of our approach achieves the best performance as well. All results indicate the superiority of the proposed framework than existing approaches.

Table 2. The results on CUHKPQ dataset.

	Comprehensive features	Facial features	Features from CNN	Total
Guo [21]	92.19%	(no result)	——	95.01%
Tang [10]	(no result)	95.21	——	97.40%
Our method	94.19%	95.58%	96.60%	97.94%

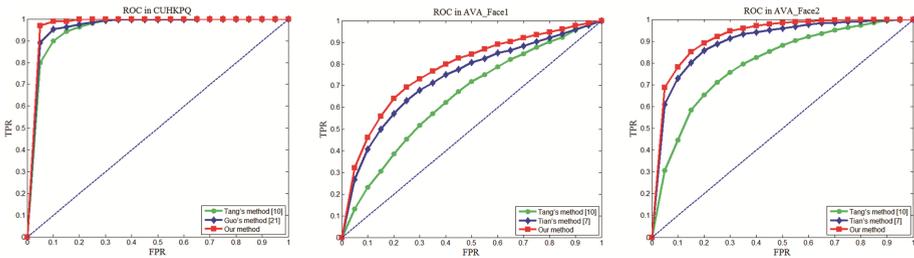


Fig. 7. The ROC curves of different approaches in (a) CUHKPQ (b) AVA_Face1 (c) AVA_Face2.

5.2.3 Experiments on the AVA_Face Dataset

In this paper, we collect labels of photos in AVA_Face in two ways so that we get AVA_Face1 and AVA_Face2 dataset. When using CNN method, we choose 80% photos of AVA_Face1 dataset as training dataset, 5% as validation dataset and 15% as testing dataset. In such a way, we are able to fine-tune our deep aesthetic model. We also fine-tune an original ResNet-50-layer [15] in this way as a baseline to prove the effectiveness for combining the feature maps of four convolution layers in our CNN model. The classification accuracies in Table 3 shows that our CNN model performs better after improving.

Table 3. The classification accuracies in AVA_Face1 by two different models.

	ResNet-50-layer	Our CNN model
AVA_Face1	79.10%	79.80%

In AVA_Face1 dataset, we extract two kinds of handcrafted features of photos in testing set and fuse the features of testing set using the CNN method by decision fusion. In AVA_Face2 dataset, all photos are tested using the best model fine-tuned by AVA_Face1. Then these features will be fused with the two kinds of handcrafted

features. To test two datasets more correctly, we using 5-cross validation in experiments. The classification accuracies of our approaches are shown in Table 4. For comparison, we also extract Tang’s features [10] and LLC features [21]. The approach in paper [7] is another baseline. The results are shown in Table 4.

We can see that our approach obtains the best performance in both of the evaluated datasets. The deep learning method [7] also receives better performance than Tang’s method [10], which is close to the performance of using our fine-tuned CNN features along. However, the fusion of comprehensive features and facial features improve the overall performance. In Fig. 7(b) and (c), the ROC curve of our approach validates the effectiveness of the proposed framework.

Table 4. The classification accuracies on AVA photos with faces dataset.

	LLC features [21]	Tang [10]	DCNN [7]	Our method
AVA_Face1	73.08%	74.28%	76.90%	80.81%
AVA_Face2	63.12%	79.22%	85.12%	86.99%

5.2.4 Analysis of Different Groups of Features

In this subsection, we evaluate the effectiveness of different groups of features on face classification using three different datasets, CUHKPQ, AVA_Face1 and AVA_Face2. The results of experiments are summarized in Table 5.

We can see that, in CUHKPQ dataset, the facial aesthetic features perform well in classifying photos with faces. When we combine the facial aesthetic features and the comprehensive features, their classification accuracy are 95.58% and 94.19%. Comparably, the accuracy of features from CNN is 96.60%. The accuracies of two kinds of handcrafted features are similar or even better than that of the features from CNN. The results suggest that facial aesthetic and comprehensive features are effective on classifying photos with faces. In particular, high quality dataset (CUHKPQ) can produce robust features and thus leads to better classification results.

However, in AVA_Face1 and AVA_Face2 dataset, the facial aesthetic features and the comprehensive features perform worse than the features from CNN. This could be due to the size and the noises of the two datasets. On the other hand, CNNs are capable of reducing the noise level and provide more robust features.

Overall, in different dataset of photos with faces, three kinds of aesthetic features have their own advantages. Also, we find that when they fused by decision fusion, we can get the best results.

Table 5. The classification accuracies of different feature groups on datasets

	CUHKPQ	AVA_Face1	AVA_Face2
Facial aesthetic features	95.58%	73.69%	76.77%
Comprehensive features	94.19%	75.65%	81.20%
Features from CNN	96.60%	79.80%	84.40%
Decision Fusion	97.94%	80.81%	86.99%

6 Conclusion

In this paper, we propose a framework for automatically assessing the aesthetic quality of photos with faces. Well-designed facial aesthetic features and features learned from CNN are extracted for this task. Then we fuse these features with comprehensive features by decision fusion and obtain the best performance compared with selected baselines on several datasets. This is different from previous researches, where the main strategy is early fusion. In addition, we also collect a new dataset of photos with faces based on AVA dataset, which is a large-scale dataset of photos with faces and has a wider application. Experiments show that our method lead to promising results. The study of aesthetic quality on photos with faces can still have many potential applications.

Acknowledgements. This work was supported by the Natural Science Foundation of Guangdong Province #2015A030313212, National Natural Science Foundation of China #U1636218, the State Scholarship Found of China #201506155081, the Science and Technology Planning project of Guangdong Province #2014B010111006 and Guangzhou Key Lab of Body Data Science #201605030011.

References

1. Wang, W., Zhao, W., et al.: An efficient image aesthetic analysis system using Hadoop. *Signal Process. Image Commun.* **39**(108), 499–508 (2015)
2. Wang, W., Cai, D., et al.: Synthesized computational aesthetic evaluation of photos. *Neurocomputing* **172**, 244–252 (2016)
3. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part III*. LNCS, vol. 3953, pp. 288–301. Springer, Heidelberg (2006). https://doi.org/10.1007/11744078_23
4. Marchesotti, L., Perronnin, F., et al.: Assessing the aesthetic quality of photographs using generic image descriptors. In: *IEEE International Conference on Computer Vision*, pp. 1784–1791 (2011)
5. Yin, W., Mei, T., Chen, C.: Assessing photo quality with geo-context and crowd sourced photos. In: *IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–6 (2012)
6. Dong, Z., Tian, X.: Multi-level photo quality assessment with multi-view features. *Neurocomputing* **168**, 308–319 (2015)
7. Dong, Z., Shen, X., Li, H., Tian, X.: Photo quality assessment with DCNN that understands image well. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) *MMM 2015, Part II*. LNCS, vol. 8936, pp. 524–535. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14442-9_57
8. Lu, X., Lin, Z., Jin, H., Yang, J., Wang, J.: Rating pictorial aesthetics using deep learning. *IEEE Trans. Multimedia* **17**(11), 2021–2034 (2015)
9. Wang, W., Yi, J., et al.: Computational Aesthetic of Image Classification and Evaluation. *J. Comput. Aided Des. Comput. Graph.* **26**(7), 1075–1083 (2014)
10. Tang, X., Luo, W., Wang, X.: Content-based photo quality assessment. *IEEE Trans. Multimedia* **15**(8), 1930–1943 (2013)
11. Li, C., Gallagher, A., Loui, A.C., et al.: Aesthetic quality assessment of consumer photos with faces. In: *IEEE International Conference on Image Processing (ICIP)*, pp. 3221–3224 (2010)

12. Murray, N., Marchesotti, L., Perronnin, F.: AVA: A large-scale database for aesthetic visual analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415 (2012)
13. Duan, H.: Research on Facial Attractiveness Analysis Based on Machine Learning. Beijing Jiaotong University (2011)
14. Jin, X., Zhao, M., Chen, X., Zhao, Q., Zhu, S.-C.: Learning artistic lighting template from portrait photographs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 101–114. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_8
15. He, K. et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
16. Seth, P., Vijaya, R.: An Optimal minimum spanning tree algorithm. *J. ACM* **49**(1), 16–34 (2002)
17. Wang, W., Cai, D., Xu, X., Liew, A.: Visual saliency detection based on region descriptors and prior knowledge. *Signal Process. Image Commun.* **29**, 424–433 (2014)
18. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255 (2009)
19. Megvii Inc. Face ++ Research Toolkit, December 2013. www.faceplusplus.com
20. Huan, R., Pan, Y.: Decision fusion strategies for SAR image target recognition. *IET Radar Sonar Navig.* **5**(7), 747–755 (2011)
21. Guo, L., et al.: Image esthetic assessment using both hand-crafting and semantic features. *Neurocomputing* **143**, 14–26 (2014)
22. Wang, W. et al.: Image aesthetic classification using parallel deep convolution neural networks. *Acta Automatica Sinica* **42**(6) (2016)
23. Xue, S., Tang, H., Tretter, D., Lin, Q.: Feature design for aesthetic inference on photos with faces. In: IEEE International Conference on Image Processing, pp. 2689–2693 (2013)
24. Yangqing, J., et al.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia (2014)