2018

Facial Attributes Analysis and Applications

Xudong Liu

Facial Attributes Analysis and Applications

Xudong Liu

Thesis submitted to the
Staler College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master's Science
in
Computer Science

Guodong Guo, Ph.D., Chair
Xin Li, Ph.D.
Shuo Wang, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2018

Keywords: Facial Attributes Recognition; Multiple Images; Attributes Inconsistency; Deep Convolutional Neural Network; Beauty Semantics; Correlation;

Abstract

Facial Attributes Analysis and Applications

Xudong Liu

Facial attributes are one of the most powerful descriptors for personality attribution. In the area of computer vision, researchers have worked on the extraction and use of attributes in face recognition. Facial attribute recognition is conventionally computed from a single image. In practice, it is quite common to capture multiple still images for each subject or to acquire a video of a subject with a number of image frames. Thus it is not rare to encounter the situation of having multiple still images or video frames of the same subject. Then it is quite natural to request a unique set of attributes about the subject given multiple face images. Naturally, how to compute the attributes given multiple images of the same subject?

Firstly, we explore whether the inconsistency exists among the attributes computed from multiple face images of the same subject. The inconsistency can be caused by the variations in images, such as the face image quality changes. Then we develop methods in view of probabilistic confidence and image quality to address the inconsistency. Experimental results show that the proposed methods can handle facial attribute estimation on either multiple still images or video frames, and can correct the incorrectly annotated labels.

In addition, by computing the correlation between the facial attributes and the beauty score, we developed an application about mining semantic descriptions from facial attributes for beauty understanding. The facial beauty description is constructed from facial attributes. This is a totally data-driven method to address facial beauty instead of the psychology study. After analyzing beauty features, we adopt these features to the original facial images for testing the beauty difference. Experimental results indicate that the beauty semantics are reasonable and beneficial for the beauty modification.

To the beloved, my parents and Yilin

Acknowledgments

I would like to thank many people for contributing to and supporting my Master's studies. It would not be possible to complete Master's degree without their contributions.

First, I would like to express my appreciation to my supervisor Prof. Guodong Guo for his guidance and support throughout my Master's study. It is his patience and expertise that have shaped me from an uncertain student into a professional researcher. This thesis would not be possible without his enormous work and help. I am very grateful for his teachings, not just about the aspects of computer vision, but for many valuable pieces of advice that will benefit me for life.

I would also like to thank Prof. Xin Li and Prof. Shuo Wang for serving as my committee members. They spent precious time and efforts on giving suggestions to my research and helping to improve and refine this thesis.

Hereby I am grateful to my lab colleagues Dr. Yu Zhu, Dr. Wentian Zhou, Dr.Guowang Mu, Linhao Li, Mohammad Iqbal Nouyed, Mohamad Al Jazaery, Yufeng Yu, Na Zhang, for their kind help and assistance during my Master's study.

Finally, I would like to thank my parents for unconditionally love and support. And a big thank to my beloved Yilin, for being always beside me!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Facial attributes are one of the most powerful descriptors for personality attribution [6]. In the area of computer vision, researchers have worked on the extraction and use of attributes in various tasks, such as object detection and classification [4, 7–10], as well as face recognition [11–14]. Facial attributes are beneficial for multiple applications, including face verification [15–17] identification [18], and face image search [19]. It is even shown that gender classification can be improved [20] by exploiting the existence of dependencies among gender, age and other facial attributes.

Facial attributes are usually computed from a single face image, e.g., [3, 15, 21, 22]. However, we are interested in a related but different problem: How to compute the attributes given multiple face images of the same subject? In other words, our interest is to extract subject-based attributes, rather than the traditional single-image-based attributes.

In practice, it is quite common to capture multiple still images for each subject or to acquire a video of a subject with a number of image frames of the subject. Thus it is not rare to encounter the situation of having multiple still images or video frames of the same subject. Then it is quite natural to request a unique set of attributes about the subject given multiple face images, which is also beneficial for face recognition.

One possible way to derive the attributes from multiple images is to compute the attributes from each image and then get one common description of the subjects. This approach may raise an issue: Is there any inconsistency among the attributes computed from the single image? And if the inconsistency exists, how to address it? All these questions will be addressed in this thesis.

On the other hand, facial attributes are good feature to describe the appearance of a person. Another interesting study in this thesis is about mining semantic descriptions from these facial attributes for beauty understanding.

Facial beauty has a large and diverse effect on human social activities, from mate choices to decision about hiring and social exchange [23]. For example, attractive people have more dates than less attractive people [24], and people are more satisfied with their dates who are more attractive [25]. Researchers also pointed out that attractive people have higher opportunities to be hired than less attractive individuals [26]. Furthermore, attractiveness even can influence judgments about the seriousness of committed crimes [27].

Over centuries, facial beauty has been an open target to psychologists, philosophers, and artists. A significant portion of studies focused on the perception of facial beauty. What is Beauty? Psychologists have studied a diverse range of factors from symmetry [28–31], averageness [32–34], sexual dimorphism [35, 36], to some other personality attributions [37].

While the studies of beauty from psychology are moving forward, computer scientists have focused on applications of facial beauty based on above findings. Researchers have designed many advanced image processing algorithms applied for beautification [1, 38–45]. The main idea is analyzing the attractive facial geometry.

Features, like shape ratio, symmetry, texture, etc. Then they train a beauty model using machine learning methods, such as SVR (Support Vector Regression), KNN (K-Nearest Neighbors) and so on. Facial beauty prediction also attracts considerable attention [46–49]. Researchers extract different features, like LBP (Local binary patterns), Gabor, AAM (Active Appearance Model), then train auto-raters using supervised learning. The beauty scores are collected by individual ratings, as labels.

These beautification and beauty prediction studies significantly rely on the concept of beauty, which is defined by psychology. In other words, computer researchers concentrate more on facial geometry features based on related psychology work. In this thesis, we propose a novel study, demonstrating the correlation between facial beauty and some specific facial attributes (e.g., Arched Eyebrows, Nose size, etc.). This study is possible due to the big data explosion in recent years. In the overview illustrated in Fig.1, we first deploy a deep convolutional neural network (CNN) to obtain a large set of the facial attribute. Then we study the correlation between the big data of attributes and two large datasets of rated beauty [1, 2]. We demonstrate both consistency and inconsistency between psychological studies arches and our results from statistical analysis. After analyzing the correlation, we believe that our attributes can also stimulate psychology study and make beauty research more convincing.

1.1 Proposal Outline

In this thesis, we are solving the problems on facial attributes estimation in multiple images, and mining semantic descriptions for beauty understanding using facial attributes. There are fives chapters in this thesis.

Deep training for facial attributes prediction is the first step for these studies. Therefore, Chapter 2 introduces deep learning for attributes prediction. It reviews the current deep learning technology for facial attributes training. We propose our deep model outperforms several works. The details of deep training are described in this Chapter 2.

The study for facial attributes in multiple images is addressed in Chapter 3. We explore the inconsistency exists among the attributes computed from multiple images of the same subject. The inconsistency can be caused by the variations in images, such as the face image quality changes. Two methods are developed to address the inconsistency. Given these methods, the unique facial attribute can be computed.

Chapter 4 discusses the semantic description for facial beauty understanding. The correlation between the beauty score and facial attribute is studied. The beauty semantics are generated by the experimental results.

Chapter 5 summarizes and concludes the discoveries made in this thesis.

1.2 Summary of Contributions

The main contributions of this dissertation are as follows:

- A new problem is proposed, i.e., computing the subject-based attributes in contrast to the traditional single-image-based. The inconsistency problem is raised when multiple face images are given.

- Two approaches are developed to address the inconsistency issue among multiple images.
- The correlation between facial beauty and distinct facial attributes is studied.
- We categorize the face attributes based on beauty level, which is beneficial for both face image beautification and medical plastic surgery.
- A new finding of approach that agrees with previous facial beauty studies especially in psychology, and we address some discrepancy between and experimental results and the psychology studies.

Chapter 2

Deep Convolutional Neural Networks for Facial Attributes Training

2.1 Introduction

Deep convolutional neural networks (CNNs) have outperformed a series of image recognition challenges [4, 8, 50, 51]. From 2012 to 2015, all winners of the ImageNet ILSRVC challenge, which consists of 1000 classes with millions images, developed improved CNNs architecture. Based on the big success of that challenge, deep CNNs have been becoming particularly popular in computer vision community. A typical deep CNNs could have the architecture : Input - Convolution - Activation Function - Pooling -FC as shown in Figure 2.1.

While facial attributes prediction in the wild is a challenging task because of the unfavorable face variations, it is critical and beneficial for face recognition. Kumar et al. [15] employed face attributes for face verification using binary classifiers trained to recognize the presence or absence of describable visual appearance (face attributes). They first extract low-level features from different regions by hand-labeling and used AdaBoost to learn the helpful features. Different features were learnt for each facial attribute, and an SVM with RBF kernel is trained as the classifier for each attribute. Their approach is timing consuming and inefficient especially in feature extraction processing. Due to the recent advances in GPUs and deep CNNs, Liu et al. [3] cascaded two CNNs, Localization Network (LNet) for

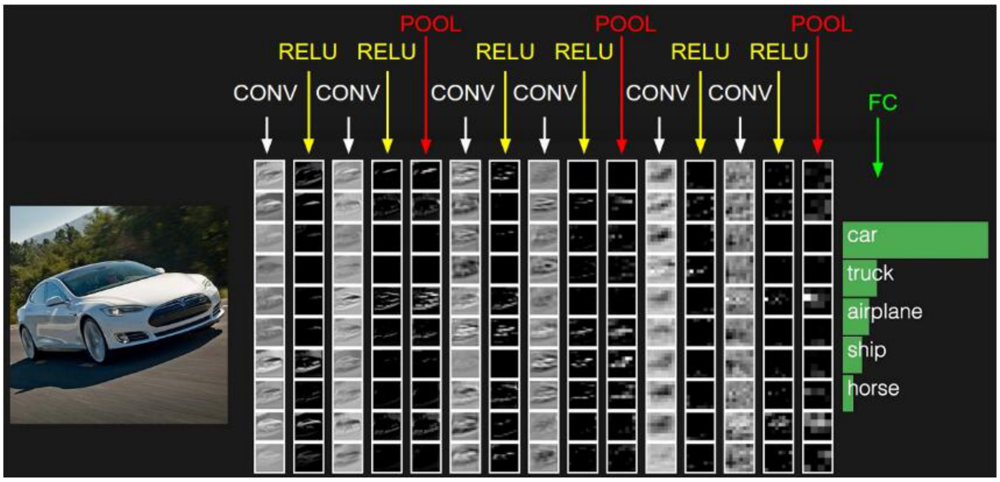


Figure 2.1: The initial volume stores the raw image pixels (left) and the last volume stores the class scores (right). Each volume of activations along the processing path is shown as a column. [52]

face localization and Attribute recognition network (ANet) for attributes prediction, which are fine-tuned jointly with attributes labels. LNet first locate the face region and ANet is trained to extract attribute features, which are fed into SVMs for final attributes prediction. They have achieved state-of-the-art performance for 40 face attributes prediction tested on CeleA and LFWA, respectively.

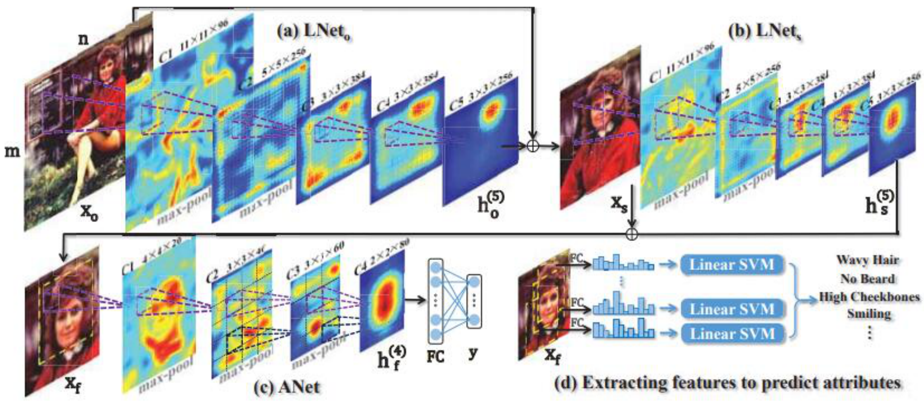


Figure 2.2: The pipeline of attribute prediction from [3]

Using CelebA [3], Zhong et al. [21] compared different features from different CNN layers and gained a better performance on face attributes prediction using the mid-level CNN feature. More recently, Rudd et al. [22] proposed a novel mixed domain adaptive optimization network (MOON) for facial attribute recognition. The architecture is developed for multi-task recognition using one single deep CNN and advances the facial attribute recognition.

2.2 Attributes Training using Deep CNNs

GoogLenet [4] achieved the state-of-the-art for classification and detection in ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The network is 22 layers deep (only counting layers with parameters) with nine Inception blocks. The proposed architecture, is a good trick for dimension reduction using 1*1 kernel and make it possible to increase the depth while saving the computational resource, named Inception as shown in Figure 2.3.

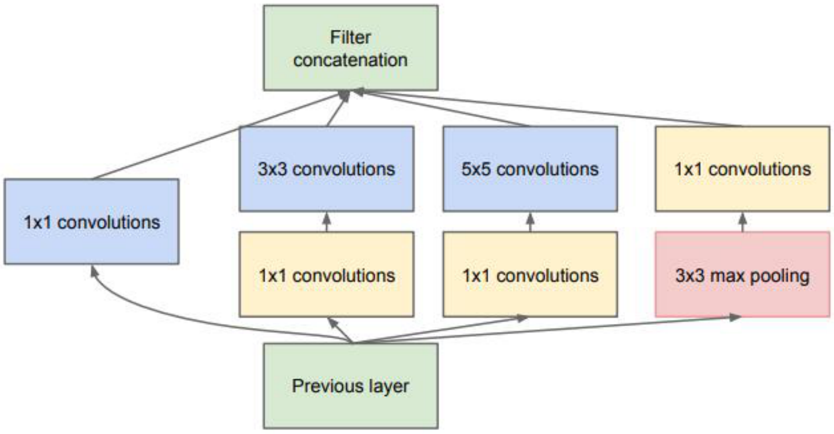


Figure 2.3: Inception module with dimensionality reduction [4].

Based on the performance on ImageNet, we deployed GoogLenet as the main architecture for attribute training. The attributes model is trained from scratch using CelebaA database. Following the protocol [3], which has three separated parts: 160,000 images of 8,000 identities are used for deep training, and the images of another 20,000 of 1,000 identities are employed to train the random forest. The remaining 20,000 images of 1,000 identities are used for testing.

2.2.1 Data Preprocessing

Before feeding images into deep CNNs, preprocessing is needed for the specific size of the face images. There are four steps for images preprocessing: face detection, landmarks detection, alignment, cropping. Constrained Local Model (CLM) [5] is used for face and landmark detection. After detection, 68 landmarks are provided as shown in Figure 2.4 . Given landmarks, the eye locations are set to [92,129] (left eye center) and [163,129] (right eye center) for alignment and then crop the images with the size of 256 *256.

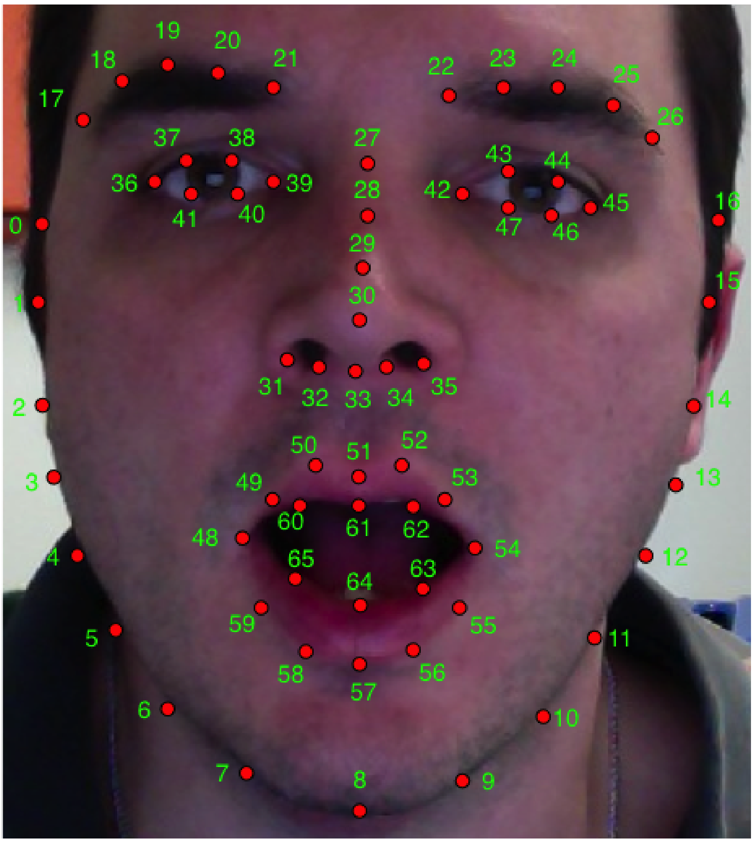


Figure 2.4: Landmark Detection using CLM [5]

2.2.2 Training Settings

As mentioned, we employ GoogLeNet [4] architecture for attributes training. For the experiment settings, the images and 40 attribute labels are stored as hdf5 files before feeding into deep CNNs using Caffe framework. Sigmoid cross-entropy is used as the loss function,

Figure 2.5: Image preprocessing: detection, alignment,cropping

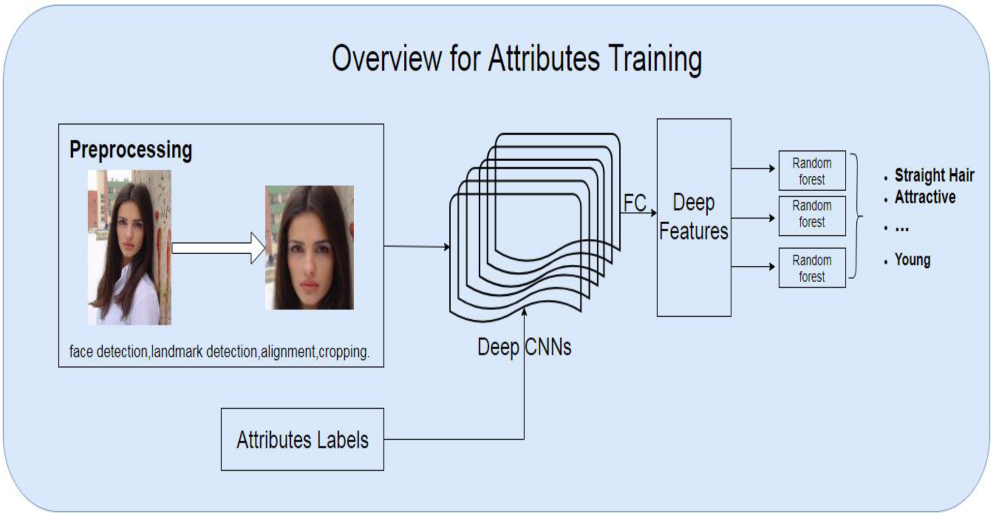


Figure 2.6: Overview for Attributes Training

the base learning rate is set 10^{-5} and reduced by a polynomial decay with gamma equals 0.5. The momentum is set to 0.9 and the weight decay equals $2 * 10^{-4}$. The number of iterations for attribute training is set to $6 * 10^5$ with the batch size of 64 using a single Titan-X GPU. Features are extracted from FC layer, and then we trained 40 random forest classifiers for attribute estimation. The overview of the training process is shown in Figure 2.6.

Loss function:

$$L = -\frac{1}{n} \sum \left[y ln a + (1 - y) \, ln \, (1 - a) \right], \tag{2.1}$$

where y denotes as the labels and a denotes as the outputs.

In addition, random forest not only can mostly avoid over-fitting compared to the single decision tree but also does not need tons of parameters to tune as SVM. For these reasons, we deploy random forest algorithm as our classifier to estimate the attributes. Random forest is much faster than SVM in our practice. After optimization of these models, we have achieved 88% accuracy over the 40 facial attributes, which is comparable to the state-of-the-art.

Table 2.1: Attributes Performance on CelebA.

Attributes	**Liu et al. [3]**	**Ours**	**Attributes**	**Liu et al. [3]**	**Ours**
5_o_Clock_Shadow	**91**	90	Black_Hair	87	**88**
No_Beard	**95**	91	Straight_Hair	73	**79**
Blond_Hair	**95**	**95**	Gray_Hair	**97**	**97**
Wavy_Hair	**80**	74	Attractive	**81**	79
Heavy_Makeup	**90**	88	Pale_Skin	91	**96**
Bags_Under_Eyes	79	**80**	Brown_Hair	80	**82**
Pointy_Nose	**72**	71	Wearing_Hat	**99**	98
Bushy_Eyebrows	90	**92**	Male	**98**	95
Wearing_Lipstick	**93**	91	Bangs	**95**	94
Mouth_Slightly_Open	**92**	75	Rosy_Cheeks	90	**93**
Big_Lips	**68**	67	Double_Chin	92	**95**
Sideburns	**96**	95	Wearing_Necktie	71	**86**
Eyeglasses	**99**	**99**	Narrow_Eyes	81	**85**
Goatee	**95**	**95**	Arched_Eyebrows	79	**81**
Oval_Face	66	**70**	Blurry	84	**95**
Wearing_Earrings	82	**84**	High_Cheekbones	**87**	83
Bald	**98**	**98**	Receding_Hairline	89	**92**
Chubby	91	**95**	Wearing_Necklace	71	**86**
Mustache	95	**96**	Big_Nose	78	**79**
Young	**87**	84	Smiling	**92**	85
Average	87	**88**			

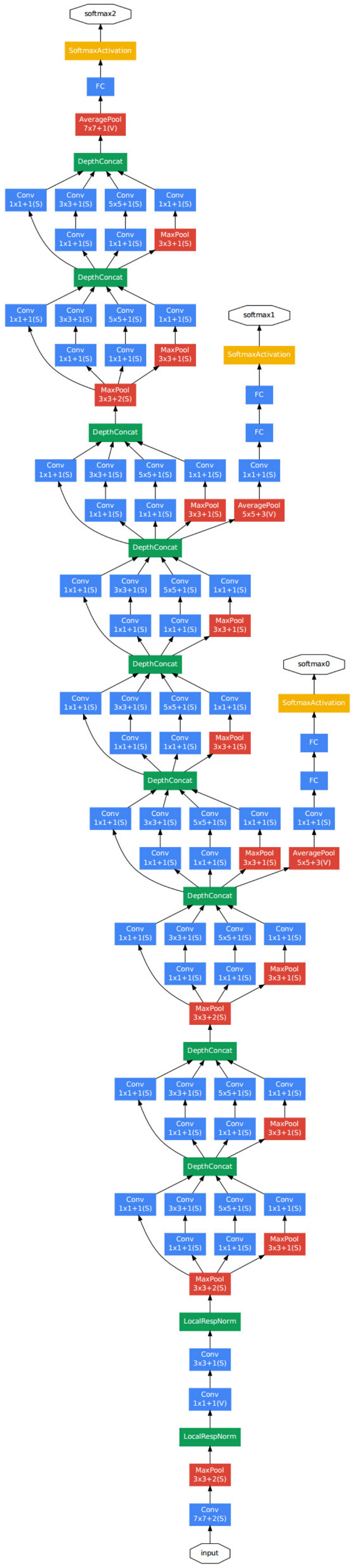


Figure 2.7: GoogLenet Architecture [4]

Chapter 3

Attributes in Multiple Facial Images

3.1 Introduction

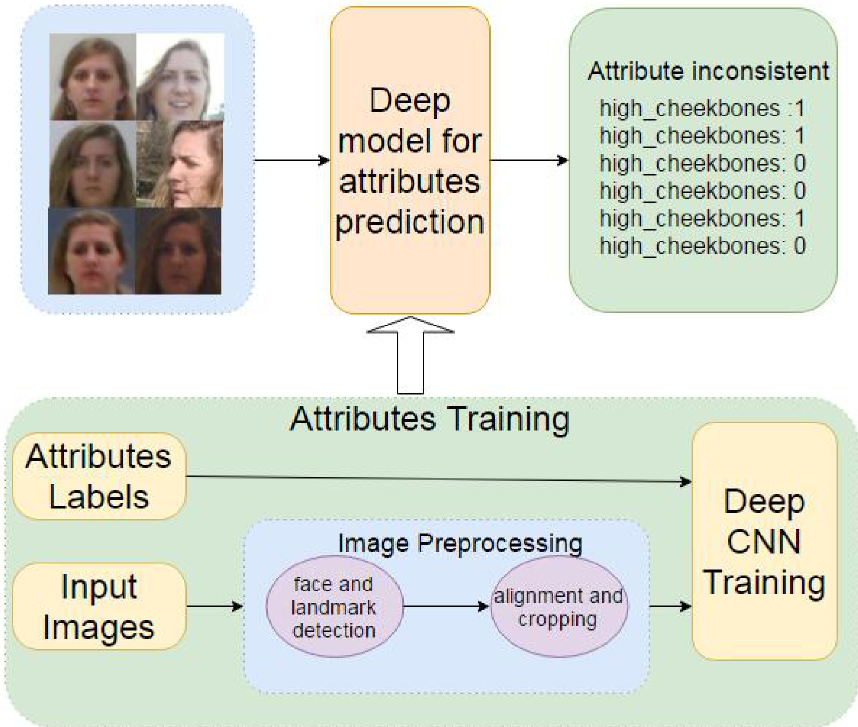


Figure 3.1: Overview of facial attributes inconsistency.

Facial attribute recognition is conventionally computed from a single image [3,15,21,22]. In practice, each subject may have multiple face images. Taking the eye size as an example, it should not change, but it may have different estimation in multiple images, which would make a negative impact on face recognition. Thus, how to compute these attributes corresponding to each subject rather than each single image is a profound work. To address this question, we explore the inconsistency issue among the attributes computed from each single image after deploying deep training for facial attributes prediction as described in Chapter 2.

Two approaches are developed to address the inconsistency issue in Section 3.3. Experimental results show that the proposed methods can handle facial attribute estimation on either multiple still images or video frames, and can correct the incorrectly annotated labels. The experiments are conducted on two large public databases with annotations of facial attributes.

3.2 Inconsistence Measure

We study the problem of face attribute inconsistency on multiple images from the same subject. Through experiments, we found that there exits inconsistency. To quantify the inconsistency, we propose to measure the inconsistency degrees, named Inconsistence Measure (IM).

Suppose there are L subjects, where $L = 1, 2, 3, \ldots$. For the l-th subject, there are N_l images, where $\sum_{l=1}^{L} N_l = N$, $N = 1, 2, 3, \ldots$. The i-th image of the l-th subject is denoted as S_l^i, where $\sum_{i=1} S_l^i = N_l$. Here we define the binary classification:

$$f_j(S_l^i) = \begin{cases} 1, & if\ \mathrm{j}^{th}\ attribute\ is\ true \\ 0, & otherwise, \end{cases} \tag{3.1}$$

where j denotes the attribute index, j=1,2,3,...,40. Then the number of positive and negative prediction results can be caculated for each attribute from each subject.

$$C_l^j(1) = \sum_{i=1}^{N_l} f_j(S_l^i). \tag{3.2}$$

$$C_l^j(0) = N_l - \sum_{i=1}^{N_l} f_j(S_l^i). \tag{3.3}$$

Accordingly, a ratio to measure the portion between the positive and negative can be computed:

$$R_l^j = \max\{C_l^j(1), C_l^j(0)\}/N_l, \tag{3.4}$$

where $R \in [0.5,1]$. If there are half positive and half negative attribute results, R equals 0.5, which means that attribute has the most inconsistent issue, whereas that attribute is consistent when R equals 1. R is a basic measure for the inconsistency issue. To have a better measure, we re-scale and re-formula the ratio, as shown in (3.5) and (3.6).

$$IM_l^j\prime = (R_l^j - 0.5)/0.5 * 100. \tag{3.5}$$

$$IM_l^j = 100 - IM_l^j\prime, \tag{3.6}$$

where $IM \in [0,100]$. The IM values can indicate the inconsistency degrees. The larger the IM, the more inconsistent the attribute. Accordingly, for the j-th attribute, IM can be calculated for all subjects:

$$IM^j = \frac{1}{L}\sum_{l=1}^{L} IM_l^j. \tag{3.7}$$

From equation (3.6), there will be no inconsistency when IM is zero. The higher IM indicates more inconsistency of an attribute. It is not difficult to understand that the attribute inconsistency will influence the face recognition performance for any attribute-based face recognition systems. For example, one person should have had the high cheekbone attribute, but it disappears because of occlusion reason during a short period of a video. Considering this problem, we propose two approaches to address the issue of attribute inconsistency in multiple images.

3.3 Methods to Address Inconsistency

To address the inconsistency issue, we develop two different approaches. The first one is based on a probabilistic confidence, and the other is to consider the image quality. Both methods combine the estimation from multiple images, and eventually, improve the attribute prediction performances at the subject level.

3.3.1 Probabilistic Confidence Criterion

Binary classifiers can be used for attribute recognition for each single image. Intuitively, an efficient classifier will not only be able to make the correct prediction but also has the highest confidence. Following this idea, we check the confidence of the result. The estimation of facial attributes trained on the CelebA achieves a comparable performance to the state-of-the-art [22](see Chapter 2), which means we have trained good deep features. Subsequently, binary classifier descriptors play an equally significant role in the final result. In this work, we deployed the random forest as the classifier.

We used 40 random forest models as the classifier descriptors. Random forest is made of plenty of decision trees. We generate each probability from these binary classifiers' outputs, denoted as $P[f_j(S_l^i) = 1]$ and $P[f_j(S_l^i) = 0]$, and define confidence as:

$$Confidence_l^{ij} = \left|P[f_j(S_l^i) = 1] - P[f_j(S_l^i) = 0]\right|, \tag{3.8}$$

then, the representation of the l-th subject for the j-th attribute is computed as following:

$$Sub_l^j = \underset{i \in N_l}{\arg\max}\, Confidence_l^{ij}. \tag{3.9}$$

As a consequence, we extract the most confident image representation for each subject. We then select the result from the highest confidence as the subject's attribute.

3.3.2 Image Quality Criterion

The face image quality may also cause the inconsistency issue in attribute recognition. We investigate 11 typical heuristic features for image quality assessment, which includes brightness [53], contrast, focus [54], illumination, illumination symmetry, sharpness, compression [55], pose estimation [56], eyes detection, mouth detection and face symmetry. We empirically assign weights to each individual measure and then add these scores to generate one final score for each image, where the weights are shown in Table 3.1. Afterwards, we select the image with the highest scores for attribute recognition for each subject.

Table 3.1: The values for image quality weight.

Feature	**Weight**	**Feature**	**Weight**
brightness	0.6	compression	0.7
contrast	0.6	pose	1.0
focus	0.8	eyes openness	0.5
illumination	1.0	mouth closeness	0.5
illumination symmetry	0.9	face symmetry	1.0
sharpness	0.8		

3.3.3 Image Fusion

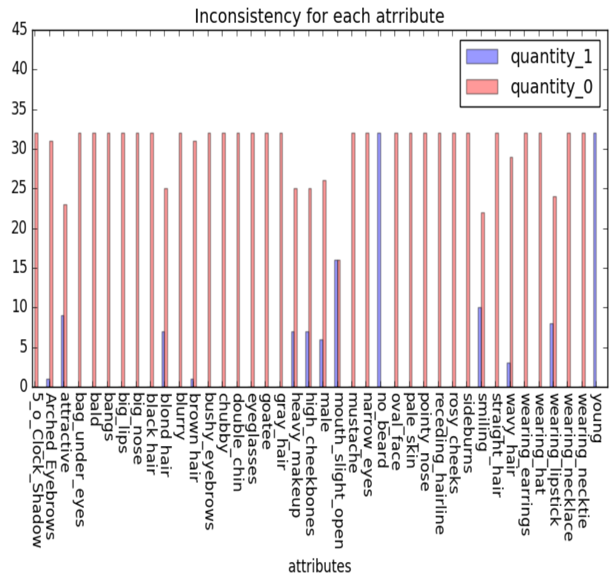
Given the above approaches, through either the probabilistic confidence or image quality criterion, we can improve performance by combining more representations. Taking probabilistic confidence as an example, we select the image that has the highest confidence. Furthermore, we select and combine the top 3 or 5 confidences for each subject. We use the majority voting as the final prediction. Eventually, the attribute recognition performance can be improved by such a fusion. The same strategy can be applied to the image quality based as well.

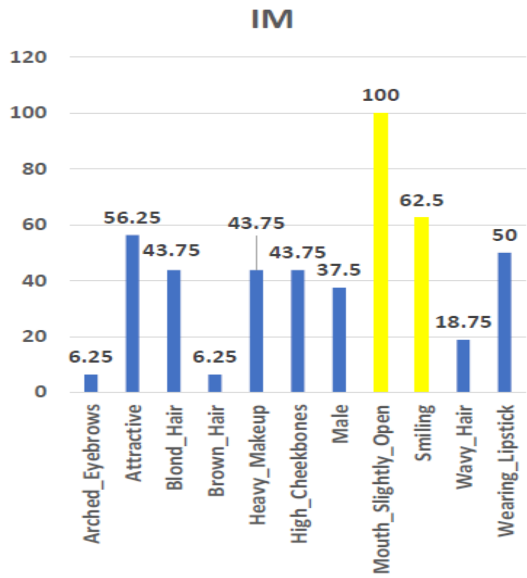
3.4 Experiment

3.4.1 Data

As discussed in Chapter 2, we employ CelebA database for face attributes training. There are 200,000 images in CelebA, including 10,000 identities, each of which contains around 20 images. For each image, 40 face attributes are labeled, in other words, 8,000,000 attributes are provided in total on this database. In order to measure the subject level facial attributes, we annotated 40 attributes on two datasets. One has 293 identities from PaSC [57]. There are 293 identities from PaSC testing dataset, including 9376 still images (about 32 images per subject), and 2802 videos (approximately ten videos per person and 100 frames per video). Another dataset is COX [58], which has 1,000 subjects, 3 videos captured for each

(a) One identity images from PaSC





(b) Attributes recognition results

(c)IM

Figure 3.2: The images show many variations in (a),(b) shows the attributes inconsistency of the subject. Facial attributes excluding attractive, mouth slight open and smiling (yellow bar in (c)) which are depended on each image. However, attributes, like Blond hair, High cheekbones, Male, which should be consistent but experimental results show inconsistent. (c) is the IM results, the higher IM the inconsistency is more serious, e.g. IM for Month_Slightly_Open is 100, which means numbers of 0 and 1 are equal.

subject with 3 different camcorders. An interactive tool for annotating facial attributes was developed, which displayed multiple face images from the same subject. A rater was asked to check each attribute. Each subject was labeled by 3 volunteers. In order to obtain the subject level labels, we finalized the labels using a majority voting to get the unique result for each attribute. Therefore, 1293 subjects with 51720 facial attributes are used in our experiments.

3.4.2 Deep Training for Facial Attributes Recognition

The details of training processing are described in Chapter 2.

3.4.3 Mutiple Still Images and Videos on PaSC

Even though both still images and videos [57] are from several locations (inside buildings and outdoors), pose angles, different distances, as well as numbers of sensors, some kind of intrinsic attributes, e.g., gender, nose size, hair color, face shape, narrow eyes, pale skin, should be consistent at least for years. In addition, many attributes, such as, arched eyebrows, bald, bangs, chubby, double chins, goatee, high cheekbones, mustache, receding hairline, sideburns, hair shape, wearing earrings, wearing necklace, wearing necktie, also should not change for each person during a short time period. However, it would be challenging for face recognition when these facial attributes become inconsistent.

For still images, i.e., several images for the same subject. We compute the IM for each subject with each of the 40 attributes, using (6). One subject example for IM is shown in Figure 3.2. We then concatenate the holistic still images using (3.7) and the whole IM are given in Table 3.2.

After IM generated, the inconsistency issue is clear in Table 3.2. We addressed the inconsistency as given in Section 3.3. As a consequence, we obtain a unique result for each attribute, and achieve 85.6% and 83.0% over 40 attributes based on the two criteria, respectively, as shown in Table 3.3.

We can also apply the strategies to video frames. The difference is that while each video is considered as a subject for the video experiments, rather, each identity is denoted as one

Table 3.2: Inconsistence Measure on PaSC.

Attributes	Still images	Video frames
Arched_Eyebrows	28.81	31.30
Attractive	5.67	5.51
Bangs	12.71	22.89
Big_Nose	0.53	0.34
Bushy_Eyebrows	0.28	0.31
Eyeglasses	63.71	60.14
Heavy_Makeup	1.98	1.36
High_Cheekbones	47.83	50.12
Male	0.19	0.38
Pointy_Nose	0.21	0.53
Straight_Hair	0.17	0.13
Wearing_Lipstick	63.52	42.50
Young	0.23	0.19

subject for still image experiments. There are several videos from the same identity in PaSC; it makes no sense if we simply combine different videos even they are from the same identity, because different videos should have their inconsistency issues. As the preceding analysis, we compute the highest confidence and the highest image quality, respectively. Afterward, we can provide unique results over 40 attributes for each video. Ultimately, the performance of videos reached 84.8% and 83.8% based on probabilistic confidence and image quality assessment, respectively, as shown in Table 3.3.

Table 3.3: Performance after selection.

	Confidence	Image Quality
PaSC Still	85.6%	83.0%
PaSC Video	84.8%	83.8%

3.4.4 Videos on COX

The COX [58] consists of 1,000 subjects and three videos for each subject. We focus on the videos, which contain several frames, and demonstrate the attribute inconsistency issue.

We first compute the inconsistency from the entire video database on COX, and the IM is calculated as shown in Table 3.4. Except for some attributes that exist for a short time, such as Mouth Slightly Open, Smiling, we are still able to find seven facial attributes that are inconsistent. As a result, we deploy our approaches to define these attributes on each video.

Table 3.4: Inconsistence Measure (IM) on COX.

Attributes	**Cam1**	**Cam2**	**Cam3**
Attractive	13.15	9.40	6.97
Bangs	3.6	0	0.68
Eyeglasses	0.34	0.02	0
High_Cheekbones	17.68	18.72	24.25
Male	0.51	0.32	0.32
Wearing_Lipstick	1.71	0.13	1.29
Young	0.32	0.56	0.02

Similar to PaSC videos, we use the binary decision confidence for each frame in each video, before the final decision. For each video, we search the most confident frame for the attribute estimation. On the other hand, there are some variations in each video clip, such as illumination, pose variation, blur, etc. As a consequence, we adopt the measured approach for image quality as we described in Section 3.3. After the quality ranking, the highest quality image frame in each video is taken as input for attribute prediction. The accuracies over 40 attributes from all three camcorders videos are shown in Figure 3.3.

3.4.5 Results from Fusion

As discussed in Section 3.3, we not only consider the best representation for each subject, but also improve the performance with fusion. From the probabilistic confidence on PaSC,

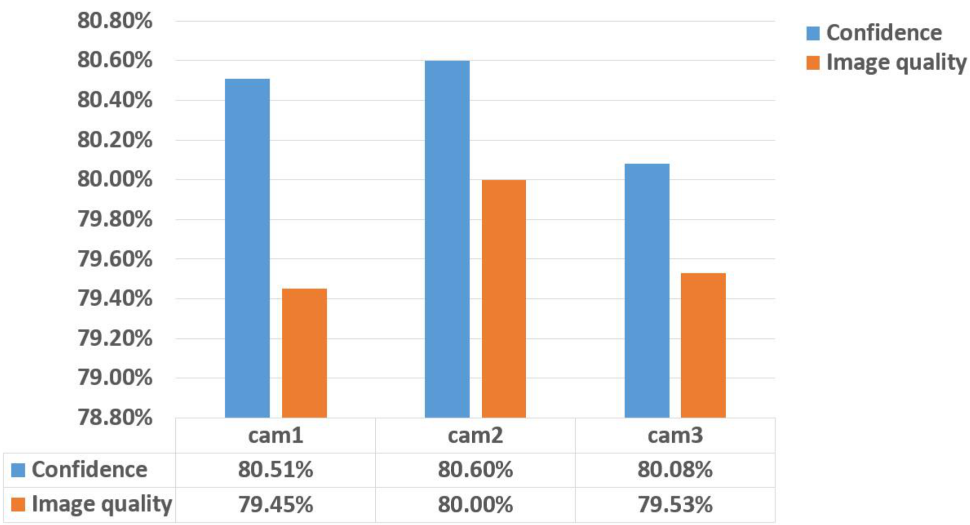


Figure 3.3: Attributes accuracy on COX

we find out the best performance (86.0%) comes up when we consider the top 3 for fusion. Additionally, given image quality, we gain the best performance with the fusion of top 5, as shown in Figure 3.4.

From the experiments, we found that it is not true to get a better result with more images to combine for probabilistic confidence criteria. When considering more images, the chance that images with weak confidence dominate the result is increasing. While the top 3 can be fused to achieve the best performance based on probabilistic confidence. For quality assessment, we can see in Figure 3.5, the images keep a high quality through top 1 to 5. Therefore, the more images we are taking, the better performance we achieve. After fusion experiments, the accuracy is improved to 86.2% both Still and videos on PaSC.

3.4.6 Correct the Incorrectly Annotated Labels on CelebA

There are 1,000 identities on CelebA testing set. We explore whether there are also inconsistencies for attributes. Similar procedures as we described on PaSC and COX datasets, we first extract the deep feature and proceed the attributes prediction based on each identity. Computed by (3.7), the IM values are generated as shown in Table 3.5.

Using our methods, we can provide unique attribute description for multiple images of the same subject. We then check whether there is also inconsistency for the attributes labels

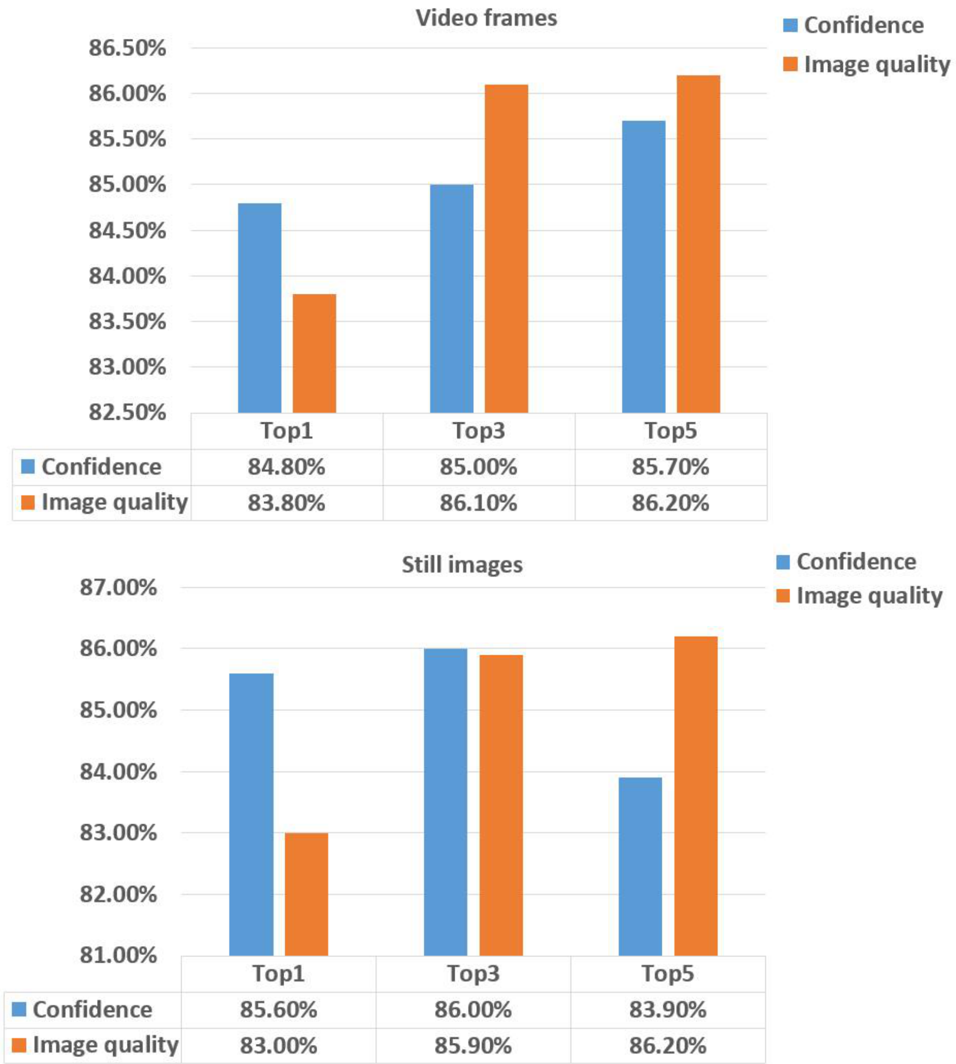


Figure 3.4: Fusion images from confidence and image quality perspective. Better performance both stills and videos on PaSC compared to top1.

(ground truth). Different from the previous procedures where the outputs are from deep features, this time we calculate their IM based on the attributes labels and the corresponding subjects. Following (3.7) , the IM is calculated for the annotated labels as shown in Table 3.6.

From Table 3.6, we can see that the ground truth labels have the inconsistency issue. Excluding those dependent attributes, Arched_Eyebrows, Pointy_Nose, and Oval_Face, etc. there is still a relatively high IM which indicates the inconsistency. Our proposed approach can handle this issue and correct the incorrectly annotated labels.

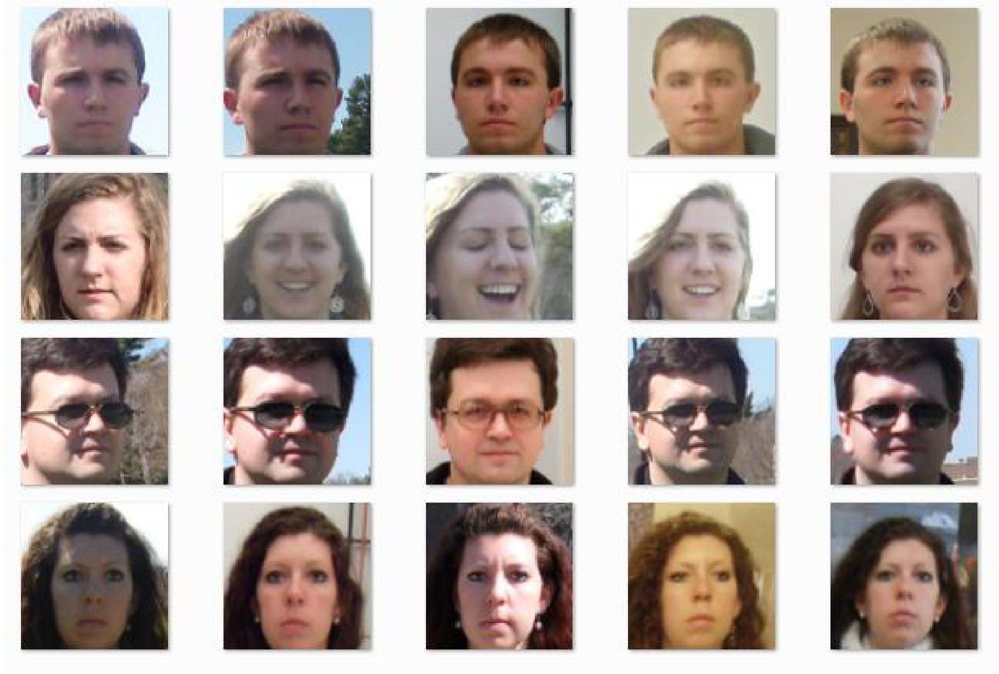


Figure 3.5: Image quality ranking examples on PaSC. top 1 to top 5 from left to right.

As we know, labeling data is expensive, but we do need these manual works to service better performance for deep learning. But how to find the label's correctness is difficult and expensive too. Taking the gender as an example, it would take massive human force to manually check the mistakes for gender annotation. Nonetheless, our method can be used to correct the errors, as shown in Figure 3.6. We can consider the highest confidence and quality or adopt the fusion idea as we described in Section 3.3, and finally provide the consistent attribute labels.

Table 3.5: Inconsistence Measure on CelebA.

Attributes	**IM**	**Attributes**	**IM**
Attractive	7.06	High_Cheekbones	24.39
Bangs	2.39	Male	6.95
Big_Nose	0.07	Mouth_Slightly_Open	65.47
Eyeglasses	7.25	Smiling	3.36
Heavy_Makeup	1.98	Wearing_Lipstick	0.54
High_Cheekbones	47.83	Young	0.49

Table 3.6: Label Inconsistence Measure on CelebA.

Attributes	**IM**	**Attributes**	**IM**	**Attributes**	**IM**
5_o_Clock_Shadow	11.31	Black_Hair	27.99	Goatee	6.42
No_Beard	11.70	Straight_Hair	29.02	Arched_Eyebrows	26.63
Blond_Hair	13.53	Gray_Hair	4.89	Oval_Face	35.24
Wavy_Hair	35.09	Attractive	31.76	Blurry	11.25
Heavy_Makeup	25.18	Pale_Skin	7.32	Wearing_Earrings	28.22
Bags_Under_Eyes	28.65	Brown_Hair	27.41	High_Cheekbones	46.46
Pointy_Nose	28.71	Wearing_Hat	7.53	Bald	2.73
Bushy_Eyebrows	16.42	Male	1.26	Receding_Hairline	13.63
Wearing_Lipstick	15.53	Bangs	18.93	Chubby	8.01
Mouth_Slightly_Open	55.50	Rosy_Cheeks	11.19	Wearing_Necklace	21.45
Big_Lips	16.93	Double_Chin	7.66	Mustache	4.77
Sideburns	6.66	Wearing_Necktie	11.69	Big_Nose	19.65
Eyeglasses	8.79	Narrow_Eyes	23.53	Smiling	52.77
Young	6.71				

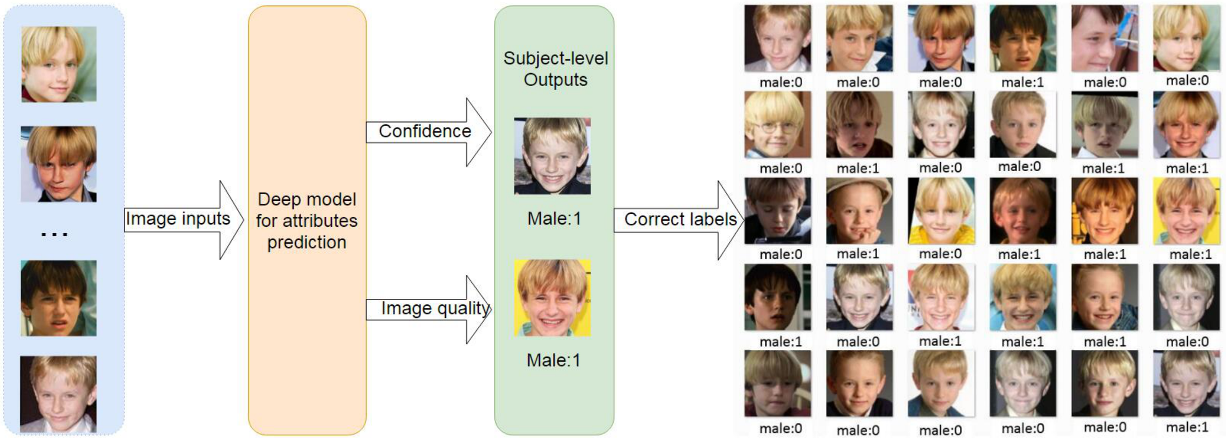


Figure 3.6: The right side shows the attribute labels of one identity on CelebA [3]. Even though they are from the same identity, the attribute label (Male) has encountered inconsistency. Using our methods, two representations are selected based on confidence and image quality, and we can output the subject-level attribute estimation to correct the incorrectly annotated labels.

3.5 Conclusion

In this Section, we proposed a novel problem to study and developed methods for facial attributes from multiple images of the same subject. We illustrated the face attributes inconsistence issue when dealing with multiple images or video frames. After that, we developed two approaches to address the problem using probabilistic confidence and image quality assessment. Given these approaches, the unique facial attribute can be computed. Moreover, the methods can be applied to correct the incorrectly annotated labels in a large database.

Chapter 4

Mining Semantic Descriptions from Data for Beauty Understanding

4.1 Introduction

Facial beauty has a large and diverse effect on human social activities, from mate choices to decision about hiring and social exchange [23]. For example, attractive people have more dates than less attractive people [24], and people are more satisfied with their dates who are more attractive [25]. In a society that is virtually obsessed by beauty, looking unpleasant or different can deeply affect self-esteem and result in social isolation, depression and serious psychological disorders [59–63]. Researchers also pointed out that attractive people have higher opportunities to be hired than less attractive individuals [26]. Furthermore, attractiveness even can influence judgments about the seriousness of committed crimes [27]. Therefore, mining the beauty semantics is a profound work.

Over centuries, facial beauty has been an open target to psychologists, philosophers, and artists. A significant portion of studies focused on the perception of facial beauty. What is Beauty? Psychologists have studied a diverse range of factors from symmetry [28–31], averageness [32–34], sexual dimorphism [35, 36], to some other personality attributions [37].

While the study of facial beauty from psychology has been conducted for centuries, the study of beauty for computer science is relatively new. However, with the widespread use of cameras and the popular social media, images are pervasive in all aspects of social life. As a

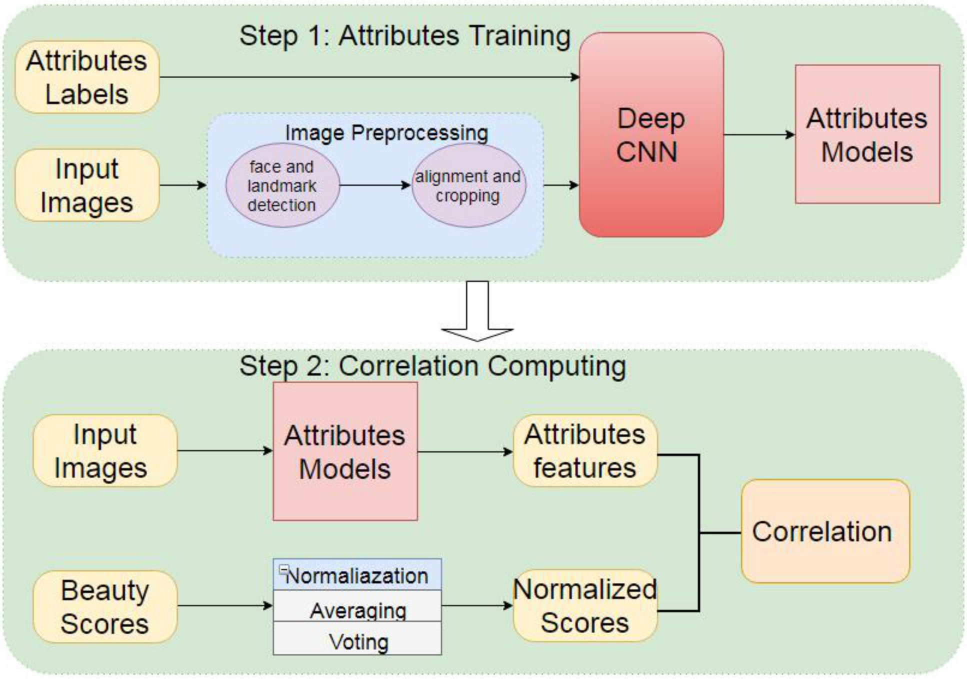


Figure 4.1: Overview of approach.

result, computer scientists have designed many advanced image processing algorithms applied for beauty enhancement (or beautification) based on psychological findings [1, 38–45]. The main idea is analyzing the attractive facial geometry features, like shape ratio, symmetry, texture, etc. Then they train a model to beautify the facial images using machine learning methods, such as SVM (Support Vector Machine) [64–66], KNN (K-Nearest Neighbors) [67] and so on. Facial beauty prediction also attracts considerable attentions [46–49, 68]. Researchers extract different features, like LBP (Local binary patterns) [69], Gabor [65, 70], AAM (Active Appearance Model), then train auto-raters using supervised learning. The beauty scores are collected by individual ratings, as labels.

The studies of beautification and beauty prediction significantly rely on the concept of beauty, which is defined by psychology. In other words, computer researchers concentrate more on facial geometry features based on related psychology work. In stead of facial geometry feature, we propose a novel study, demonstrating the correlation between facial beauty and high-level facial features (e.g., Arched Eyebrows, Nose size, etc.). Authors in [38] proposed that high-level facial features play a critical role in beauty estimation, which motivates us to figure out what high-level features are behind of beauty. This study is driven by the big data explosion as well as the promising performance of deep learning. As the overview illustrated in Figure 4.1, a deep convolutional neural network (CNN) is deployed for facial attributes training. Then we study the correlation between the massive high-level features and two large data sets of rated beauty [1, 2]. After computing the correlation, facial attributes are categorized based on beauty level. Finally, we demonstrate both consistency and inconsistency between psychological studies arches and our findings from statistical analysis. After analyzing the correlation, our findings not only can contribute to psychology study and make the beauty research more convincing but also are beneficial for beauty enhancement.

The major contributions are fourfold:

- Facial attributes in two data sets are estimated using deep neural networks.
- We first present the correlation between facial beauty and distinct facial attributes.
- We categorize the face attributes based on beauty level, which is beneficial for both face image beautification and medical plastic surgery.
- The approach presents new findings that agrees with previous facial beauty studies especially in psychology, and we address some discrepancy between and experimental results and the psychology studies.

This Chapter is organized as follows. An investigation into facial beauty is demonstrated in Section 4.2. Section 4.3 describes our proposed approach. Section 4.4 details the experiments. The further analysis is discussed in Section 4.5. Finally, conclusions and future perspective are presented in Section .

4.2 Investigation

In this section, we investigate and organize the current studies on beauty from the view of both psychology and computer science.

4.2.1 Psychology Research on Facial Attractiveness (Beauty)

What is beauty? The question has tackled by the psychologist as well as philosophers for centuries. The well-known saying Beauty is in the eye of the beholder indicates that individual beauty is subjective and unpredictable because of our knowledge of a persons particular environment and culture. However, across many studies, it has been found that there is a high cross-culture agreement in attractiveness rating of the face [71–73]. People everywhere are using similar criteria in their judgment, and there are some universal features about attractiveness, in other words, the perception of facial beauty is not decided by particular people but is a global standard.

On this stage, we investigate the beauty factors from the psychological perspective, such as symmetry, averageness, and sexual dimorphism. Psychologists have addressed that symmetry has a positive influence on attractiveness [30, 74]. Moreover, Galton et al. [32] first noted that multiple faces blended together are more attractive than the constituent faces, which indicates the averaging face is another positive influence on attractiveness. Additionally, several studies have documented people show a preference for feminine looking faces no matter what is the actual gender of the face [75–79]. Moreover, some personality traits are also reported to affect attractiveness. For example, faces shown with smiles rated as more attractive and as having more positive personality traits than neutral faces [80]. Although psychologists have analyzed these aspects, exactly what specific features make a face beautiful remains poorly defined. How to validate the standard of beauty based on data science is still unsolved. The question and missing parts will be addressed in this chapter.

4.2.2 Computational Attractiveness Research

The study of beauty in psychology has been undertaken for centuries, but this study started in the field of computer science just a few decades ago. However, the fact that beauty plays such a pivotal role in society has encouraged computer scientists to research this field, which results in various applications springing up recently, especially mobile Apps.

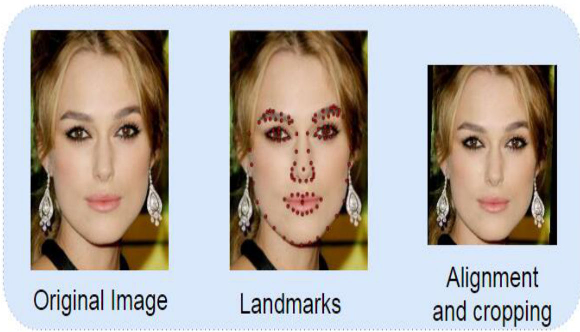
Numbers of researchers have demonstrated their contribution on how to beautify the still images and to predict facial beauty. Chen et al. [1] proposed a hypothesis on facial beauty perception. They found out the weighted averages of two geometric features is better and adopt their hypotheses on beautification model using SVR and have achieved the state-of-the-art geometric based face beautification. Liu et.al. [81] presented a purely landmark-based, data-driven method to compute three kinds of geometric facial features for a 2.5D hybrid attractiveness computational model. Lu et.al. [82] prosed a facial skin beautification framework to remove facial spots based on layer dictionary learning and sparse representation. Leyvan et al. [38] focus on enhancing the attractiveness of human faces in frontal view. They present face warping towards the beauty-weighted average of the k closer samples in face space. They also proposed that a small local adjustment may result in an appreciable impact on the facial attractiveness (partly enhance). It inspires us to find out which parts are related to attractiveness so that we can just decorate these small pieces to beautify, for example, eyes, instead of the whole face. Authors in [83] also address that high-level features are beneficial to beauty prediction, which further motivates us to figure out which specific attributes affect the beauty. The finding of the high-level features related to face beauty furthermore can contribute to the beauty enhancement (or beautification).

From our investigation, it is evident that a large, high-quality dataset is required for facial beauty research, but the biggest scale of the current public dataset is only thousand-s. Establishing a large facial beauty dataset with variety has always been recognized as a difficulty both financially and technically. The major challenge is obtaining human ratings for large-scale images. Even though studies [71–73] indicate that there is a high degree of agreement in facial beauty, there still exists inconsistence when rating attractiveness scores from different individuals. The larger the score in scales are, the more obvious is the inconsistency, and more time is consumed during ratings. The only way to reduce or eliminate this issue and to obtain a convincing beauty dataset is to involve in more human rater and apply averaging or voting for each image. That is because even though we are able to build large face dataset via Internet, the lack of rating scores for sufficient variable data set is a bottleneck. Authors in [84] proposed that most beauty rating systems, as well as those attempting to beautify faces, rely on samples of faces rated at different levels of attractiveness for training or comparisons. Therefore, large beauty face databases should be constructed. Thanks to [1, 2], we finally obtain a relatively large data sets to progress our work.

4.3 Approach

As shown in Fig.1 overview of approach: Data preprocessing, attributes training, feature extraction, correlation calculation are four major steps. All these procedures will be addressed in this section.

4.3.1 Data Preprocessing



(a) One example for image preprocessing

Attributes	result	Attributes	result	Attributes	result
5_o_Clock_Shadow	0	Black_Hair	0	Goatee	0
No_Beard	1	Straight_Hair	0	Arched_Eyebrows	0
Blond_Hair	1	Gray_Hair	0	Oval_Face	0
Wavy_Hair	0	Attractive	1	Blurry	0
Heavy_Makeup	1	Pale_Skin	0	Wearing_Earrings	0
Bags_Under_Eyes	0	Brown_Hair	0	High_Cheekbones	1
Pointy_Nose	0	Wearing_Hat	0	Bald	0
Bushy_Eyebrows	0	Male	0	Receding_Hairline	0
Wearing_Lipstick	1	Bangs	0	Chubby	0
Mouth_Slightly_Open	0	Rosy_Cheeks	0	Wearing_Necklace	0
Big_Lips	0	Double_Chin	0	Mustache	0
Sideburns	0	Wearing_Necktie	0	Big_Nose	0
Eyeglasses	0	Narrow_Eyes	0	Smiling	1
Young	1				

(b) Attributes recognition results

Figure 4.2: One example for image preprocessing and the corresponding attributes results

Before feeding images into deep CNNs, preprocessing is needed for the specific size of the face images. There are four steps for images preprocessing: face detection, landmarks detection, alignment, cropping. Constrained Local Model (CLM) [5] is used for face and landmark detection. After detection, 68 landmarks are provided as shown in Figure 4.2 . Given landmarks, the eye locations are set to [92,129] (left eye center) and [163,129] (right eye center) for alignment and then crop the images with the size of 256 *256.

In addition to image preprocessing, beauty scores also need normalization because there are some inconsistencies when multiple people rate per image and we adopt majority voting and averaging methods to generate scores from [1, 2], respectively.

4.3.2 Attributes Training

In this paper, we employ GoogLeNet [4] architecture for attributes training. The network is 22 layers deep (only counting layers with parameters) with nine Inception blocks.

The overview of the training process is shown in Figure 2.6. First, images and attribute labels are fed into the deep CNNs, features are extracted from Fully Connected layer (FC). Then 40 random forest classifiers are trained for attributes prediction and final output the attributes results. More details are described in Chapter 2

4.3.3 Correlation Calculation

After we generated the 40 face attributes as well as the normalized beauty score, our objective is to study the correlations between attributes and beauty. Pearson correlation coefficient is computed for correlation measurement.

Pearson correlation coefficient r is used to investigate the relationship between two variables. It is calculated by:

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{4.1}$$

$$\rho_{X,Y} = \frac{cov\left(X, Y\right)}{\rho_X \rho_Y} \tag{4.2}$$

Where cov is the covariance, ρ_X and ρ_Y are the standard deviation of X and Y, respectively. The values of Pearson correlation coefficient is between -1 and 1. If $r > 0$, it indicates a positive correlation between X and Y, if $r < 0$, it indicates a negative correlation. When $r = 0$, it indicates no correlation between X and Y [85]. The objective is to analysis how these specific attributes influence human beauty after calculating the correlation.

4.4 Experiment

4.4.1 Data

For beauty analysis, we employ two rated data sets for experimental analysis. Chen et al. [1] built a beauty database with diversified and ethnic groups (we refer to Beauty

Table 4.1: Beauty Data sets Description

Database	**Size**	**Ethnicity**	**Gender**	**Scale**	**Rators**	**Normalization**
Beauty 799 [1]	799	diverse	Only Female	3	25	Voting
The 10k US [2]	2222	Caucasian	Female and Male	5	12	Averaging

799). They collected 799 female face images in total, 390 celebrity face images including Miss Universe, Miss World, movie stars, and super models, and 409 common face images. They use a 3-point integer scale for rating: 3 for unattractive, 2 for common, and 1 for very attractive. Each image is rated by 25 volunteers. Another data set is the 10k US Adult Face Database [2], which consists of 10168 American adults, 2222 faces are labeled on Amazon Mechanical Turk with 12 respondents. Different from rating on Beauty 799 [1], the 10k US Adult Face [26] use a 5-point integer attractiveness scale, 5 represents the most attractive, 1 is for most unattractive. Details see in TABLE 4.1.

For attributes training, Liu et al. [3] published the labeled CelebA for face attribute prediction. There are around 200,000 images containing 10,000 identities, each of which has 40 attributes labels. Following the protocol [3], which has three separated parts: 160,000 images of 8,000 identities are used for deep training, and the images of another 20,000 of 1,000 identities are employed to train the random forest. The remaining 20,000 images of 1,000 identities are used for testing.

4.4.2 Facial Attributes Training Settings

Details see in Chapter 2.

4.4.3 Correlation Between Attributes and Beauty

After obtaining facial attributes, the correlation between these attributes and beauty scores are computed. Since each attribute is a binary decision, the variable for attribute is either 1 or 0, and the other variable is beauty score. There are some attributes differences between females and males related to beauty. The objective is mining the beauty semantics for both females and males. As a consequence, the experiments are divided into three parts, one is for the female, another is for the male and an experiment for both females and males. Since Beauty 799 [1] only consists of females, we only need to take the whole data set and conclude the results of the female analysis. However, we experimentally separate female and male from computing the correlation on the 10k US [2]. Details on correlation computing are described in the following part.

Details on correlation computing are illustrated in this part. Let X_i^j denote as the $i-th$ attribute in the $j-th$ image, n is size of the data set, i is from 1 to 40. Y_j is the beauty score of the $j-th$ image. Applied by equation (1),

$$r_i = \frac{\sum_{j=1}^{n}\left(X_i^j - \bar{X}_i\right)\left(Y_j - \bar{Y}\right)}{\sqrt{\sum_{j=1}^{n}\left(X_i^j - \bar{X}_i\right)^2}\sqrt{\sum_{j=1}^{n}\left(Y_j - \bar{Y}\right)^2}} \tag{4.3}$$

Therefore, we are able to compute all correlations including 40 different attributes as shown in Table 2,3,4 and Fig.5. Some attributes are not able to calculate the correlation because all of these attributes are predicted the same value 0 or 1, then the standard deviation equals zero. Correlation tells you about how one particular variable varies when the other variable moves around, however, the standard deviation equals zero implying that variable is not moving around. As a result, there is no correlation between these two variables. Taking the double chin attribute as an example, we cannot define the relationship between beauty and this feature when all images do not consist of the double chin attribute. Another concern is that our correlations are not very large, that is because, for each attribute, the output is a binary classifier, the range of variables is poor. To demonstrate this concern, we adopt a chi-squared test to address there indeed exists relations even if the Pearson correlation is small. In other words, this correlation is not employed as the classic interpretation, weak correlation still works in this study.

4.4.4 Changing Attributes for Testing Beauty Difference

After correlation computed, additional experiments are made to test the difference when modifying the facial attributes.

In this experiment, we employ a Deep Feature Interpolation [86] for changing face attributes, and then we manually rank the beauty from these changed images and correspond-

Table 4.2: Positive Attributes for Attractiveness on Beauty 799

Attribute	r	**Attribute**	r
Arched Eyebrow	-0.110	Heavy Makeup	-0.203
High Cheekbone	-0.107	No Beard	-0.040
Pale Skin	-0.010	Pointy Nose	-0.010
Wavy Hair	-0.062	Wearing Earring	-0.047
Wearing Lipstick	-0.245	Young	-0.088
Straight Hair	-0.010		

Table 4.3: Negative Attributes for Attractiveness on Beauty 799

Attribute	r	**Attribute**	r
Big Nose	0.054	Black Hair	0.062
Blond Hair	0.073	Bushy Eyebrows	0.034
Gray Hair	0.034	Male	0.206
Mouth Slightly Open	0.086	Smiling	0.005

ing to their original images. There are two major steps for this interpolation. First, we generate two lists from CelebA, one consists of the attributes which have positive effects on beauty based on correlation results, another does not include these. Then we subtract these two categories as a difference and feed it into VGG [7] architecture with a target image, whose attributes are going to be changed. Applying the correlation results, we select 50 images both from Beauty 799 and the 10k US, we then tune different parameters in Deep Feature Interpolation and generate the corresponding images to the originals as shown in Fig.4. In this experiment, we take high cheekbones as an example, the experimental results show that the cheekbone is becoming higher with different parameters and epochs. However, the image quality is another concern with iterations increasing. Therefore, the most challenging work is tuning ideal parameters, and the objective is not only changing attributes images but the images can keep relatively high quality as well because the image quality also plays a significant role in beauty judgment. As a consequence, this is a tradeoff based on this tool [86], heavier modification with lower image quality is shown in Fig4. Thus, the target

Table 4.4: Beauty Semantics on 10K US

Attributes for Attractiveness	**Attributes for Unattractiveness**
Black Hair	Big Nose
Blond Hair	Bushy Eyebrows
Heavy Makeup	Male
High Cheekbone	Mouth Slightly Open
No Beard	Straight Hair
Sideburns	Young
Smiling	
Wearing Lipstick	

image with a mildly modified is hand-picked for rating. As we mentioned, there 100 images in total for ranking. Seven graduate students involved in this ranking work, each volunteer is asked to pick the most beautiful image from two options, one is original, and the other is the modified one. After that, we statistically compute the ratio between the actual changing and the original one as shown in Table 5. We can conclude that our correlation work well even this tool has some limitations.

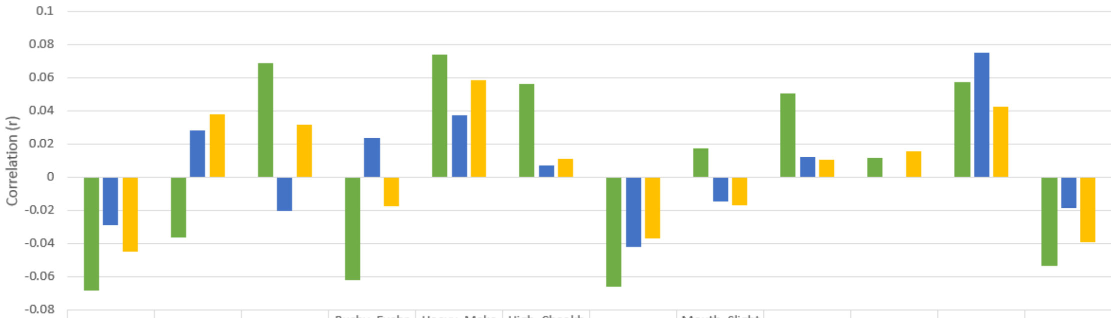
4.5 Experimental analysis

4.5.1 Mining Beauty Semantics on Beauty 799 [1]

As we previously mentioned, Beauty 799 data set only consist of females, and the rating scores (Y) are 1 (very attractive), 2 (common), and 3 (unattractive). Applied above Pearson

Table 4.5: Identical Beauty Semantics on both Data Sets [1, 2]

Attributes for Attractiveness	**Attributes for Unattractiveness**
Heavy Makeup	Big Nose
High Cheekbone	Bushy Eyebrows
No Beard	Male
Wearing Lipstick	Mouth Slightly Open



	Big_Nose	Black_Hair	Blond_Hair	Bushy_Eyebrows	Heavy_Makeup	High_Cheekbones	Male	Mouth_Slightly_Open	No_Beard	Sideburns	Smiling	Straight_Hair
Female	-0.068603921	-0.036226091	0.068736638	-0.062079726	0.073862517	0.056070655	-0.066311132	0.017305272	0.05073072	0.011598152	0.057127171	-0.053466828
Male	-0.028899563	0.02840701	-0.020202908	0.023504564	0.037525288	0.006965162	-0.042242607	-0.014769976	0.012103437		0.074950833	-0.018401683
Both	-0.044829239	0.037906038	0.031646615	-0.017737056	0.0586835	0.011307545	-0.036989398	-0.016844296	0.010624835	0.015514544	0.042543011	-0.039369777

Figure 4.3: . Not only considering the whole database, the studies of female ,male are illustrated, respectively

correlation interpretation, take Arched Eyebrows as an example, its r equals -0.11 which indicates Arched Eyebrows has a negative correlation with beauty score (Y). Since Arched Eyebrows only can be chosen 0 or 1, specifically, it indicates when people have the attribute of Arched Eyebrows (1), the beauty score (Y) is going down, but small beauty score (Y) represents more attractive (from original rating). Therefore, the attributes with negative r have a positive impact on beauty. As a result, as shown in Table 2, we are able to generate all the correlations between face attributes and beauty degree on Beauty 799 [1].

From Beauty 799 data set, first, we can conclude that people who have such attributes, like, Arched Eyebrows, Makeup, High Cheekbone, No Beard, Pale Skin, Pointy Nose, Straight Hair, Wavy Hair, Wearing Earrings, Wearing Lipstick, Young, are more attractive. On the other hand, it is recognized as less attractive when people have these attributes, such as Big Nose, Black Hair, Blond Hair, Bushy Eyebrows, Gray Hair, Male, Mouth Open, Smiling. However, it should be pointed out that there are some inconsistencies when a person has all these attributes. For example, usually an attractive person is not able to show Straight Hair and Wavy Hair simultaneously, but it indicates that both Straight Hair and Wavy Hair have a positive influence on attractiveness in this data set.

4.5.2 Mining Beauty Semantics on 10k US [2]

Different from Beauty 799, the 10k US Adult Face Database [2] contains more images and consists of both males and females but only Americans. The scales of beauty score in 10k US Adult Face Database [2] are five levels, and 1 indicates the least attractive, 5 indicates the most attractive. After attribute features extracting, the correlation between beauty score and attribute feature is computed by Pearson Correlation as shown in Fig. 4.3, and positive correlation suggests people with these attributes have a positive impact on beauty in this scenario.

As previously mentioned, we divide three parts for analyzing the beauty semantics on the 10k US [2]. When considering the whole data set including both female and male (see in TABLE 4.4), the attributes with Black Hair, Heavy Makeup, High Cheekbone, No Beard, Smiling and Wearing Lipstick are positive to a persons beauty. On the other hand, these attributes including Big Nose, Blond Hair, Bushy Eyebrows, Male, Mouth Slightly Open, Straight Hair as well as Young have a negative impact on beauty. That is the general beauty semantics conclusion on the 10k US.

More specifically, when we experimentally study the beauty semantics only using female face images, Blond Hair and Sideburns are considered as the positive effect on beauty besides the general results as we discussed above. On the other hand, apart from the general negative attributes, the attributes with Black Hair and Bushy Eyebrows for female are negative to beauty. When studying the male beauty, we find out all those attributes which would increase beauty still have a positive effect on beauty except Blond Hair, instead, Blond Hair is considered as an unattractive attribute for male's scenario. Also, the attributes which are negative for beauty are same as the whole data sets results.

4.5.3 Feminine Features for Beauty

Not only are we able to conclude the objective beauty semantics using data statistics, but there is another interesting finding that feminine features are recognized as more attractive compared to masculine features. From psychological perspective, there are considerable evidences that feminine features increase the attractiveness of male and female faces across different cultures [75–79]. In our experiment, the attributes that Makeup, No Beard, Wearing Earrings, Wearing Lipsticks are the naturally feminine feature, at least most males would not adopt them. Therefore, it is a consistent interpretation that these attributes have a positive effect on attractiveness both from our statistical results and psychology. Besides, there is a gender attribute named Male of which the prediction is reliable tested in CelebA from our deep model (95% accuracy). However, we found an interesting result that some females are estimated as males from the prediction model in Beauty 799 database, which indicates those females have some masculine features (Male tendency) that confused the machine. Furthermore, this Male bias attribute decreases the attractiveness from the correlation analysis. That is a contrary evidence that turns out feminine features increase the attractiveness based on our finding.

4.5.4 Inconsistent and Identical Beauty Semantics

As aforementioned, there are some intrinsic differences between these two databases [1,2]. As a result, the semantical results have some inconsistencies. Some interesting findings are illustrated: the US adults have a preference on Black Hair and Blond Hair but not on Straight Hair, which turns out an opposite conclusion to the results from Beauty 799. This phenomenon might be affected by environment, different culture might have some slight preference for hair color and shape. Apart from the inconsistency crossing the inter-database, in 10k US, we point out that Black Hair and Bushy Eyebrows are considered as attractive attributes referring to the male results. However, it is an absolute reverse when it comes to female results, both Black Hair and Bushy Eyebrows have a negative effect on beauty understanding. Another inconsistent attribute is Blond Hair between females and males, for females it is recognized as a positive attribute on beauty, but it is negative for males.

Even some inconsistencies occur in [1, 2], there still exists some identical semantics for both positive and negative on attractiveness in [1, 2]. The attributes that identically play a positive or negative role in beauty from these two relatively large data sets are summarized in TABLE 4.5. For example, these attributes: Heavy Makeup, High Cheekbones, No Beard, Wearing Lipstick would increase attractiveness (Beauty). However, the attributes with Big

Nose, Bushy Eyebrows, Male bias (refer to the female), Mouth Open have a negative impact on attractiveness.

Additionally, these results are conducted by large data statistics, and there are some opposite conclusions between the studies of psychology and our experiments. Smiling is considered as an attractiveness attribute from psychology research [80], but we regard smiling as an unattractiveness attribute from Beauty 799 [1]. Another failure is 'young,' that is unreasonable to be recognized an unattractiveness attribute from the 10k US Adult Face [2]. Both 'smiling' and 'young' have been shown there is an inconsistency between these two data sets. That is what we need to continue to deal with in the future.

4.6 Conclusion

In this paper, firstly, we address how beauty affects our social outcomes. We then investigate the current studies on facial beauty both from psychology and computer science points of view and point out their weakness. Our novelty is mining semantic descriptions for beauty understanding by computing correlation between beauty and specific attributes. Our study not only can provide the data-driven evidence for psychological beauty studies but more significantly, reveals the high-level features for beauty semantics which are beneficial for beauty enhancement. Although there are some inconsistencies between the two data sets [1,2] due to the data variations, many identical semantics are categorized based on beauty level. This paper is a novel to guide further studies of beauty using the big data.

Chapter 5

Summary

This final chapter is to summarize the conclusions of this thesis.

Chapter 2 proposes the deep training for facial attributes. It briefly introduces the success of deep CNNs in computer vision. The training processing including image processing and deep training settings are discussed.

Chapter 3 presents a novel question : how to address facial attributes in multiple images. Two methods are developed when subject-based attributes inconsistence occurs. Experimental results show that the proposed methods not only can handle multiple images but also can correct the incorrectly annotated labels.

Chapter 4 focuses on mining semantic descriptions from attributes for beauty understanding. The correlation between facial attributes and beauty scores on two different datasets is studied. Experimental results show that the beauty semantics are categorized by facial attributes and correlate the psychology study.

References

[1] David Zhang, Fangmei Chen, and Yong Xu, "A new hypothesis on facial beauty perception," in *Computer Models for Facial Beauty Analysis*, pp. 143–163. Springer, 2016.

[2] Wilma A Bainbridge, Phillip Isola, and Aude Oliva, "The intrinsic memorability of face photographs.," *Journal of Experimental Psychology: General*, vol. 142, no. 4, pp. 1323, 2013.

[3] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3730–3738.

[4] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[5] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. IEEE, 2013, pp. 354–361.

[6] Ian S Penton-Voak, Nicholas Pound, Anthony C Little, and David I Perrett, "Personality judgments from natural and composite facial images: More evidence for a kernel of truth in social perception," *Social Cognition*, vol. 24, no. 5, pp. 607–640, 2006.

[7] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] Xingyu Zeng, Wanli Ouyang, Bin Yang, Junjie Yan, and Xiaogang Wang, "Gated bi-directional cnn for object detection," in *European Conference on Computer Vision*. Springer, 2016, pp. 354–369.

[10] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[11] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[12] Yi Sun, Ding Liang, Xiaogang Wang, and Xiaoou Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[13] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.

[14] Erjin Zhou, Zhimin Cao, and Qi Yin, "Naive-deep face recognition: Touching the limit of lfw benchmark or not?," *arXiv preprint arXiv:1501.04690*, 2015.

[15] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 365–372.

[16] Fengyi Song, Xiaoyang Tan, and Songcan Chen, "Exploiting relationship between attributes for improved face verification," *Computer Vision and Image Understanding*, vol. 122, pp. 143–154, 2014.

[17] Thomas Berg and Peter N Belhumeur, "Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 955–962.

[18] Ohil K Manyam, Neeraj Kumar, Peter Belhumeur, and David Kriegman, "Two faces are better than one: Face recognition in group photographs," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–8.

[19] Yu-Heng Lei, Yan-Ying Chen, Lime Iida, Bor-Chun Chen, Hsiao-Hang Su, and Winston H Hsu, "Photo search by face positions and facial attributes on touch devices," in *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 2011, pp. 651–654.

[20] Juan Bekios-Calfa, José M Buenaposada, and Luis Baumela, "Robust gender recognition by exploiting facial attributes dependencies," *Pattern Recognition Letters*, vol. 36, pp. 228–234, 2014.

[21] Yang Zhong, Josephine Sullivan, and Haibo Li, "Leveraging mid-level deep representations for predicting face attributes in the wild," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3239–3243.

[22] Ethan M Rudd, Manuel Günther, and Terrance E Boult, “Moon: A mixed objective optimization network for the recognition of facial attributes,” in *European Conference on Computer Vision*. Springer, 2016, pp. 19–35.

[23] Anthony C Little, Benedict C Jones, and Lisa M DeBruine, “Facial attractiveness: evolutionary based research,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1571, pp. 1638–1659, 2011.

[24] Ronald E Riggio and Stanley B Woll, “The role of nonverbal cues and physical attractiveness in the selection of dating partners,” *Journal of Social and Personal Relationships*, vol. 1, no. 3, pp. 347–357, 1984.

[25] Ellen Berscheid, Karen Dion, Elaine Walster, and G William Walster, “Physical attractiveness and dating choice: A test of the matching hypothesis,” *Journal of experimental social psychology*, vol. 7, no. 2, pp. 173–189, 1971.

[26] Thomas F Cash and Robert N Kilcullen, “The aye of the beholder: Susceptibility to sexism and beautyism in the evaluation of managerial applicants,” *Journal of Applied Social Psychology*, vol. 15, no. 4, pp. 591–605, 1985.

[27] Harold Sigall and Nancy Ostrove, “Beautiful but dangerous: effects of offender attractiveness and nature of the crime on juridic judgment.,” *Journal of Personality and Social Psychology*, vol. 31, no. 3, pp. 410, 1975.

[28] Randy Thornhill and Steven W Gangestad, “Human fluctuating asymmetry and sexual behavior,” *Psychological Science*, vol. 5, no. 5, pp. 297–302, 1994.

[29] Linda Mealey, Ruth Bridgstock, and Grant C Townsend, “Symmetry and perceived facial attractiveness: a monozygotic co-twin comparison.,” *Journal of personality and social psychology*, vol. 76, no. 1, pp. 151, 1999.

[30] Gillian Rhodes, Fiona Proffitt, Jonathon M Grady, and Alex Sumich, “Facial symmetry and the perception of beauty,” *Psychonomic Bulletin & Review*, vol. 5, no. 4, pp. 659–669, 1998.

[31] Joanna E Scheib, Steven W Gangestad, and Randy Thornhill, “Facial attractiveness, symmetry and cues of good genes,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 266, no. 1431, pp. 1913–1917, 1999.

[32] FJ Galton, “Composite portraits,” *Nature*, vol. 18, pp. 97–100, 1878.

[33] Judith H Langlois and Lori A Roggman, “Attractive faces are only average,” *Psychological science*, vol. 1, no. 2, pp. 115–121, 1990.

[34] Gillian Rhodes and Tanya Tremewan, “Averageness, exaggeration, and facial attractiveness,” *Psychological science*, vol. 7, no. 2, pp. 105–110, 1996.

[35] Gillian Rhodes, Janelle Chan, Leslie A Zebrowitz, and Leigh W Simmons, “Does sexual dimorphism in human faces signal health?,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 270, no. Suppl 1, pp. S93–S95, 2003.

[36] Ian S Penton-Voak and Jennie Y Chen, “High salivary testosterone is linked to masculine male facial appearance in humans,” *Evolution and Human Behavior*, vol. 25, no. 4, pp. 229–241, 2004.

[37] Anthony C Little, D Michael Burt, and David I Perrett, “What is good is beautiful: Face preference reflects desired personality,” *Personality and Individual Differences*, vol. 41, no. 6, pp. 1107–1118, 2006.

[38] Tommer Leyvand, Daniel Cohen-Or, Gideon Dror, and Dani Lischinski, “Data-driven enhancement of facial attractiveness,” *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 38, 2008.

[39] Kaoru Arakawa and Kohei Nomoto, “A system for beautifying face images using interactive evolutionary computing,” in *Intelligent Signal Processing and Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium on.* IEEE, 2005, pp. 9–12.

[40] Stefano Melacci, Lorenzo Sarti, Marco Maggini, and Marco Gori, “A template-based approach to automatic face enhancement,” *Pattern Analysis and Applications*, vol. 13, no. 3, pp. 289–300, 2010.

[41] Kristina Scherbaum, Tobias Ritschel, Matthias Hullin, Thorsten Thormählen, Volker Blanz, and Hans-Peter Seidel, “Computer-suggested facial makeup,” in *Computer Graphics Forum.* Wiley Online Library, 2011, vol. 30, pp. 485–492.

[42] Qiqi Liao, Xiaogang Jin, and Wenting Zeng, “Enhancing the symmetry and proportion of 3d face geometry,” *IEEE transactions on visualization and computer graphics*, vol. 18, no. 10, pp. 1704–1716, 2012.

[43] Lei Zhang, David Zhang, Ming-Ming Sun, and Fang-Mei Chen, “Facial beauty analysis based on geometric feature: Toward attractiveness assessment application,” *Expert Systems with Applications*, vol. 82, pp. 252–265, 2017.

[44] Shu Liu, Yang-Yu Fan, Ashok Samal, and Zhe Guo, “Advances in computational facial attractiveness methods,” *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16633–16663, 2016.

[45] David Zhang, Fangmei Chen, and Yong Xu, “Beauty analysis fusion model of texture and geometric features,” in *Computer Models for Facial Beauty Analysis*, pp. 89–101. Springer, 2016.

[46] Jintu Fan, KP Chau, Xianfu Wan, Lili Zhai, and Ethan Lau, “Prediction of facial attractiveness from facial proportions,” *Pattern Recognition*, vol. 45, no. 6, pp. 2326–2334, 2012.

[47] Douglas Gray, Kai Yu, Wei Xu, and Yihong Gong, “Predicting facial beauty without landmarks,” in *European Conference on Computer Vision.* Springer, 2010, pp. 434–447.

[48] Kendra Schmid, David Marx, and Ashok Samal, "Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios," *Pattern Recognition*, vol. 41, no. 8, pp. 2710–2717, 2008.

[49] Tam V Nguyen, Si Liu, Bingbing Ni, Jun Tan, Yong Rui, and Shuicheng Yan, "Towards decrypting attractiveness via multi-modality cues," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, no. 4, pp. 28, 2013.

[50] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[51] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[52] Fei-Fei Li, "cs231n.stanford," `http://cs231n.stanford.edu/`, Accessed April 4, 2017.

[53] Mohammad A Haque, Kamal Nasrollahi, and Thomas B Moeslund, "Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*. IEEE, 2013, pp. 443–448.

[54] Ayman Abaza, Mary Ann Harrison, Thirimachos Bourlai, and Arun Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, vol. 3, no. 4, pp. 314–324, 2014.

[55] Zhou Wang, Hamid R Sheikh, and Alan C Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*. IEEE, 2002, vol. 1, pp. I–I.

[56] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2879–2886.

[57] J Ross Beveridge, P Jonathon Phillips, David S Bolme, Bruce A Draper, Geof H Givens, Yui Man Lui, Mohammad Nayeem Teli, Hao Zhang, W Todd Scruggs, Kevin W Bowyer, et al., "The challenge of face recognition from digital point-and-shoot cameras," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[58] Zhiwu Huang, Shiguang Shan, Ruiping Wang, Haihong Zhang, Shihong Lao, Alifu Kuerban, and Xilin Chen, "A benchmark and comparative study of video-based face recognition on cox face database," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5967–5981, 2015.

[59] Ray Bull and Nichola Rumsey, *The social psychology of facial appearance*, Springer Science & Business Media, 2012.

[60] Marlene Rankin, Gregory L Borah, Arthur W Perry, and Philip D Wey, "Quality-of-life outcomes after cosmetic surgery.," *Plastic and reconstructive surgery*, vol. 102, no. 6, pp. 2139–45, 1998.

[61] Katharine A Phillips, Susan L McElroy, Paul E Keck Jr, Harrison G Pope Jr, and James I Hudson, "Body dysmorphic disorder: 30 cases of imagined ugliness," *The American Journal of Psychiatry*, vol. 150, no. 2, pp. 302, 1993.

[62] Frances Cooke Macgregor, "Social, psychological and cultural dimensions of cosmetic and reconstructive plastic surgery," *Aesthetic plastic surgery*, vol. 13, no. 1, pp. 1–8, 1989.

[63] Eileen Bradbury, "The psychology of aesthetic plastic surgery," *Aesthetic plastic surgery*, vol. 18, no. 3, pp. 301–305, 1994.

[64] Amit Kagian, Gideon Dror, Tommer Leyvand, Daniel Cohen-Or, and Eytan Ruppin, "A humanlike predictor of facial attractiveness," in *Advances in Neural Information Processing Systems*, 2007, pp. 649–656.

[65] Yili Chen, Huiyun Mao, and Lianwen Jin, "A novel method for evaluating facial attractiveness," in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*. IEEE, 2010, pp. 1382–1386.

[66] Huiyun Mao, Lianwen Jin, and Minghui Du, "Automatic classification of chinese female facial beauty using support vector machine," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, 2009, pp. 4842–4846.

[67] Parham Aarabi, Dominic Hughes, Keyvan Mohajer, and Majid Emami, "The automatic measurement of facial beauty," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*. IEEE, 2001, vol. 4, pp. 2644–2647.

[68] Hani Altwaijry and Serge Belongie, "Relative ranking of facial attractiveness," in *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. IEEE, 2013, pp. 117–124.

[69] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[70] Jacob Whitehill and Javier R Movellan, "Personalized facial attractiveness prediction," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. IEEE, 2008, pp. 1–7.

[71] Yael Eisenthal, Gideon Dror, and Eytan Ruppin, "Facial attractiveness: Beauty and the machine," *Neural Computation*, vol. 18, no. 1, pp. 119–142, 2006.

[72] Judith H Langlois, Lisa Kalakanis, Adam J Rubenstein, Andrea Larson, Monica Hallam, and Monica Smoot, "Maxims or myths of beauty? a meta-analytic and theoretical review.," *Psychological bulletin*, vol. 126, no. 3, pp. 390, 2000.

[73] Michael R Cunningham, Alan R Roberts, Anita P Barbee, Perri B Druen, and Cheng-Huan Wu, “” their ideas of beauty are, on the whole, the same as ours”: Consistency and variability in the cross-cultural perception of female physical attractiveness.,” *Journal of Personality and Social Psychology*, vol. 68, no. 2, pp. 261, 1995.

[74] David I Perrett, D Michael Burt, Ian S Penton-Voak, Kieran J Lee, Duncan A Rowland, and Rachel Edwards, “Symmetry and human facial attractiveness,” *Evolution and human behavior*, vol. 20, no. 5, pp. 295–307, 1999.

[75] Anthony C Little and Peter JB Hancock, “The role of masculinity and distinctiveness in judgments of human male facial attractiveness,” *British Journal of Psychology*, vol. 93, no. 4, pp. 451–464, 2002.

[76] David I Perrett, Kieran J Lee, Ian Penton-Voak, D Rowland, Sakiko Yoshikawa, D Michael Burt, SP Henzi, Duncan L Castles, and Shigeru Akamatsu, “Effects of sexual dimorphism on facial attractiveness,” *Nature*, vol. 394, no. 6696, pp. 884, 1998.

[77] Gillian Rhodes, Catherine Hickford, and Linda Jeffery, “Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive?,” *British journal of psychology*, vol. 91, no. 1, pp. 125–140, 2000.

[78] Anthony C Little, D Michael Burt, Ian S Penton-Voak, and David I Perrett, “Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 268, no. 1462, pp. 39–44, 2001.

[79] Anthony C Little, Benedict C Jones, Ian S Penton-Voak, D Michael Burt, and David I Perrett, “Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 269, no. 1496, pp. 1095–1100, 2002.

[80] Ran Hassin and Yaacov Trope, “Facing faces: studies on the cognitive aspects of physiognomy.,” *Journal of personality and social psychology*, vol. 78, no. 5, pp. 837, 2000.

[81] Shu Liu, Yang-Yu Fan, Zhe Guo, Ashok Samal, and Afan Ali, “A landmark-based data-driven approach on 2.5 d facial attractiveness computation,” *Neurocomputing*, vol. 238, pp. 168–178, 2017.

[82] Xi Lu, Xiaobin Chang, Xiaohua Xie, Jian-Fang Hu, and Wei-Shi Zheng, “Facial skin beautification via sparse representation over learned layer dictionary,” in *Neural Networks (IJCNN), 2016 International Joint Conference on*. IEEE, 2016, pp. 2534–2539.

[83] Fangmei Chen, Xihua Xiao, and David Zhang, “Data-driven facial beauty analysis: Prediction, retrieval and manipulation,” *IEEE Transactions on Affective Computing*, 2016.

[84] Aldo Laurentini and Andrea Bottino, “Computer analysis of face beauty: A survey,” *Computer Vision and Image Understanding*, vol. 125, pp. 184–199, 2014.

[85] James W Cooley and John W Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation*, vol. 19, no. 90, pp. 297–301, 1965.

[86] Paul Upchurch, Jacob Gardner, Kavita Bala, Robert Pless, Noah Snavely, and Kilian Q Weinberger, "Deep feature interpolation for image content changes," *arXiv preprint arXiv:1611.05507*, 2016.