

## Attributes in Multiple Facial Images

Xudong Liu and Guodong Guo

Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA

**Abstract**—Facial attribute recognition is conventionally computed from a single image. 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 deploy deep training for facial attributes prediction, and we explore the inconsistency issue among the attributes computed from each single image. Then, we develop two approaches to address the inconsistency issue. 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.

### I. INTRODUCTION

Facial attributes are one of the most powerful descriptors for personality attribution [1]. 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 [2], [3], [4], [5], [6], as well as face recognition [7], [8], [9], [10]. Facial attributes are beneficial for multiple applications, including face verification [11], [12], [13] identification [14], and face image search [15]. It is even shown that gender classification can be improved [16] 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., [11], [17], [18], [19]. 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?

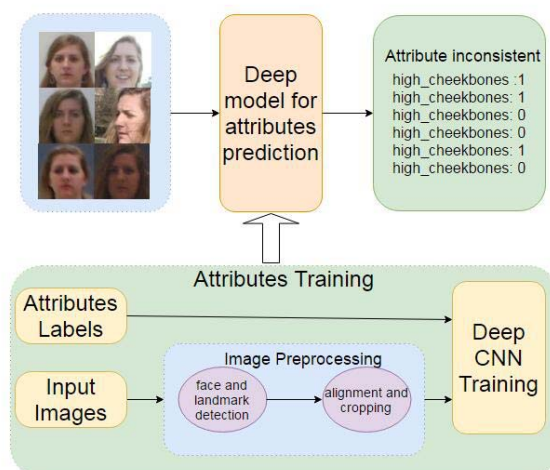


Fig. 1. Overview of attributes inconsistency.

In this paper, 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 to address the inconsistency.

Our main contributions include:

- We present a new problem, 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.
- Annotations of 40 attributes for two databases with a number of still images and video frames.
- Correct the incorrectly annotated attribute labels.

### II. RELATED WORK

Kumar et al. [11] employed face attributes for face verification using binary classifiers trained to recognize the presence or absence of describable visual appearance (face attributes). Due to the recent advances in GPUs and deep learning, Liu et al. [17] cascaded two CNNs, LNet for face localization and ANet for attributes prediction, which are fine-tuned jointly with attributes labels. They have achieved state-of-the-art performance for 40 face attributes prediction tested on CeleA and LFWA, respectively. Using [17], Zhong et al. [18] 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. [19] proposed a novel mixed domain adaptive optimization network (MOON) for facial attribute recognition. Almost all existing works focus on the estimation of face attributes from a single image. In contrast, a somewhat similar but different problem: how to compute the attributes given multiple images of the same subject? The multiple images can come from still images or video frames. When the attributes are computed from each single image, is there any inconsistency among them? If yes, how to address the inconsistency? All these questions will be addressed in the following.

### III. INCONSISTENCE MEASURE

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

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

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

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

$$C_l^j(1) = \sum_{i=1}^{N_l} f_j(S_l^i). \quad (2)$$

$$C_l^j(0) = N_l - \sum_{i=1}^{N_l} f_j(S_l^i). \quad (3)$$

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

$$R_l^j = \max\{C_l^j(1), C_l^j(0)\}/N_l, \quad (4)$$

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 (5) and (6).

$$IM_l^j = (R_l^j - 0.5)/0.5 * 100. \quad (5)$$

$$IM_l^j = 100 - IM_l^j, \quad (6)$$

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:

$$IM^j = \frac{1}{L} \sum_{l=1}^L IM_l^j. \quad (7)$$

From equation (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.

### IV. ADDRESS THE 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.

#### A. 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 [19](see Section V), 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:

$$Confidence_l^{ij} = |P[f_j(S_l^i) = 1] - P[f_j(S_l^i) = 0]|, \quad (8)$$

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

$$Sub_l^j = \arg \max_{i \in N_l} Confidence_l^{ij}. \quad (9)$$

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.

### B. 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 [20], contrast, focus [21], illumination, illumination symmetry, sharpness, compression [22], pose estimation [23], 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 I. Afterwards, we select the image with the highest scores for attribute recognition for each subject.

TABLE I  
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		

### C. 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.

## V. EXPERIMENT

### A. Data

In this work, 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 [24]. 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 [25], which has 1,000 subjects, 3 videos captured for each 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.

### B. Deep Training for Facial Attributes Recognition

Liu et al. [17], released the labeled CelebA to the public and they reached 87% accuracy over 40 attributes using LNet+ANet. Zhong et al. [18] proposed to leverage the mid-level representations from off-the-shelf architecture to tackle the attribute prediction problem for faces in the wild. They deployed different deep architecture, but both of them construct SVM as attributes classifier. Authors in [19] proposed a novel mixed objective optimization network (MOON) to handle the imbalanced data and advanced state of the art in facial attributes recognition.

We deploy the GoogLeNet [2] architecture for training deep model and random forest for the classifier. Sigmoid cross-entropy is used as the loss function, the learning rate is  $10^{-4}$  with a polynomial decay. Features are extracted from FC layer, and then we trained 40 random forest classifiers for attribute estimation using the deep features. Following the protocol [17], 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.

In addition, random forest not only can mostly avoid overfitting 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 87.7% accuracy over the 40 facial attributes, which is comparable to the state-of-the-art.

### C. Multiple Still Images and Videos on PaSC

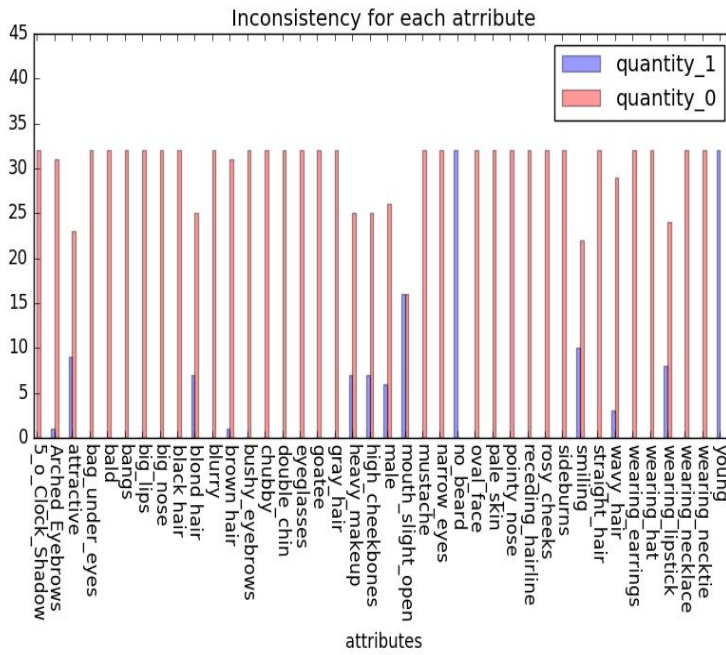
Even though both still images and videos [24] are from several locations (inside buildings and outdoors), pose angles, different distances, as well as numbers of sensors, some kind of intrinsic attributes, e.g., gender, nose size, hair color, face shape, narrow eyes, pale skin, should be consistent at 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 Fig. 2. We then concatenate the holistic still images using (7) and the whole IM are given in TABLE II.

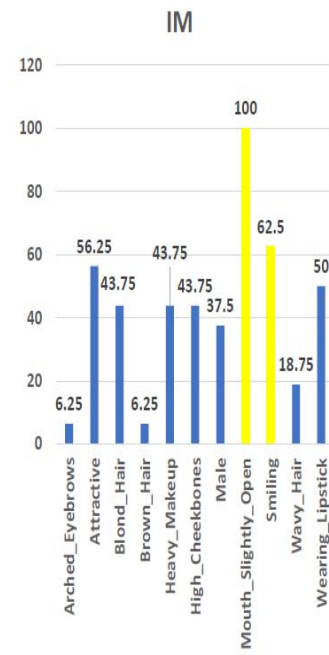
After IM generated, the inconsistency issue is clear in TABLE II. We addressed the inconsistency as given in Section IV. As a consequence, we obtain a unique result for each attribute, and achieve 85.6% and 83.0% over 40



(a)



(b)



(c)

Fig. 2. (a) Examples of one identity from PaSC still images and corresponding to the 40 attributes recognition results (b) as well as the generated IM (c). The images show many variations in (a), 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.(best viewed in color)

attributes based on the two criteria, respectively, as shown in TABLE III.

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 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 III.

#### D. Videos on COX

The COX [25] consists of 1,000 subjects and three videos for each subject. We focus on the videos, which contain

TABLE II  
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

TABLE III  
PERFORMANCE AFTER SELECTION.

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

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 IV. 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 IV  
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 IV. 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 Fig.3.

#### E. Results from Fusion

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

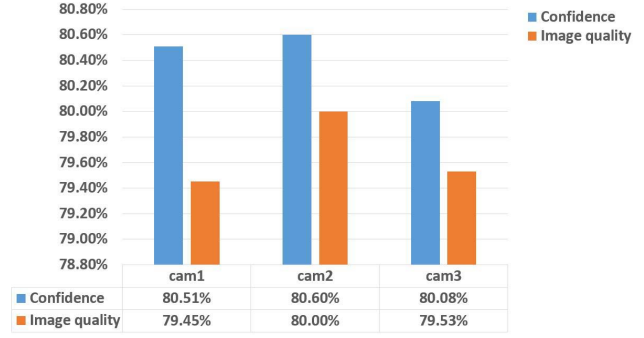


Fig. 3. Attributes accuracy on COX

on PaSC, 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 Fig.4.

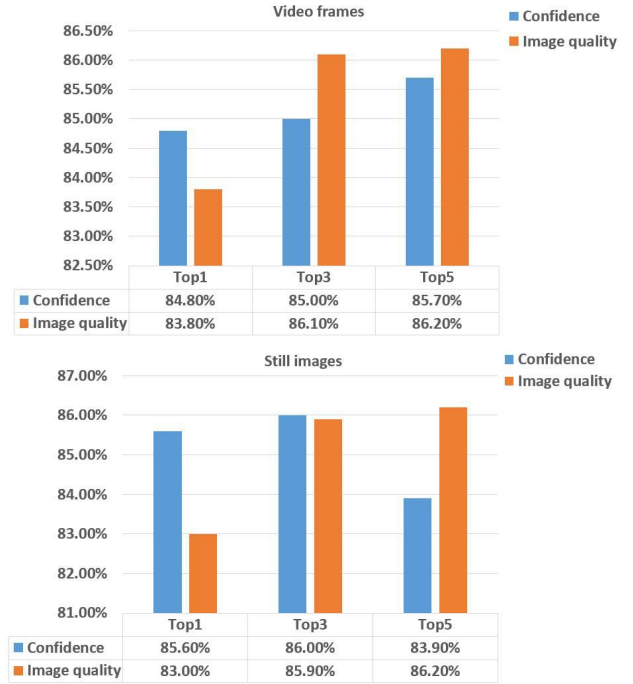


Fig. 4. Fusion images from confidence and image quality perspective. Better performance both stills and videos on PaSC compared to top1.

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 Fig.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



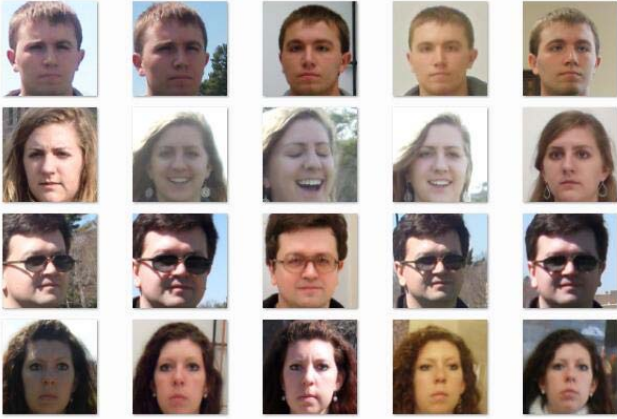


Fig. 5. Image quality ranking examples on PaSC. top 1 to top 5 from left to right.

and videos on PaSC.

#### F. 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 (7), the IM values are generated as shown in TABLE V.

TABLE V  
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

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 (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 (7), the IM is calculated for the annotated labels as shown in TABLE VI.

From TABLE VI, 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.

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 Fig.6. We can consider the highest confidence and quality or adopt the fusion idea as we described in Section IV, and finally provide the consistent attribute labels.

## VI. CONCLUSION

In this work, 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 inconsistency 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, our methods can be applied to correct the incorrectly annotated labels in a large database.

## VII. ACKNOWLEDGMENTS

This work is partly supported by a NSF-CITeR grant and a WV HEPC grant.

## REFERENCES

- [1] I. S. Penton-Voak, N. Pound, A. C. Little, and D. 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.
- [2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [5] X. Zeng, W. Ouyang, B. Yang, J. Yan, and X. Wang, "Gated bi-directional cnn for object detection," in *European Conference on Computer Vision*, pp. 354–369, Springer, 2016.
- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.
- [7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014.
- [8] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.
- [9] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.
- [10] E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of lfw benchmark or not?," *arXiv preprint arXiv:1501.04690*, 2015.
- [11] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 365–372, IEEE, 2009.
- [12] F. Song, X. Tan, and S. Chen, "Exploiting relationship between attributes for improved face verification," *Computer Vision and Image Understanding*, vol. 122, pp. 143–154, 2014.
- [13] T. Berg and P. 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*, pp. 955–962, 2013.

TABLE VI  
LABEL INCONSISTENCY MEASURE ON CELEBA.

Attributes	IM	Attributes	IM	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

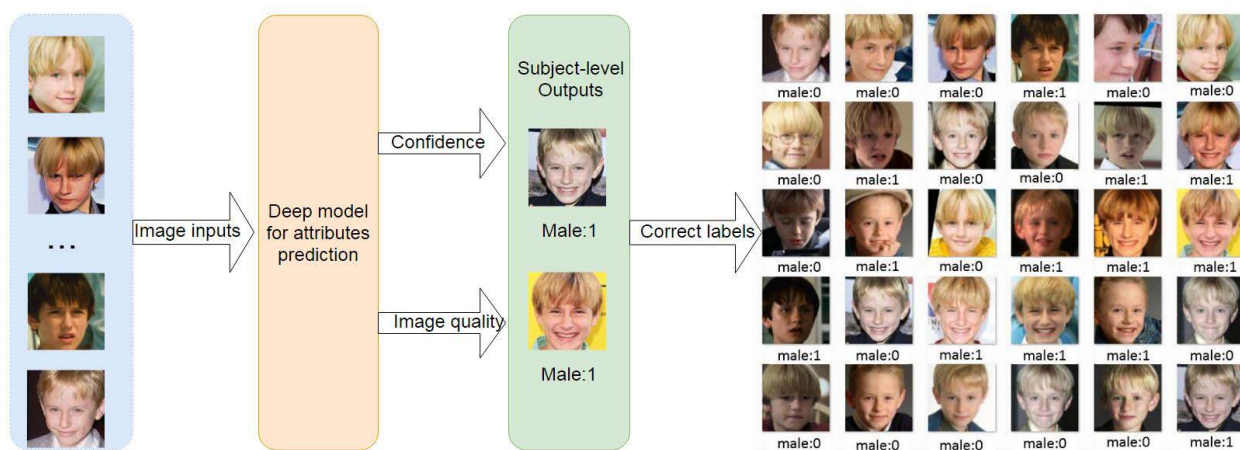


Fig. 6. The right side shows the attribute labels of one identity on CelebA [17]. Even though they are from the same identity, the attribute label (Male) has encountered inconsistency. Using our methods, two representations are selected based on confidence and image quality, and we can output the subject-level attribute estimation to correct the incorrectly annotated labels.

- [14] O. K. Manyam, N. Kumar, P. Belhumeur, and D. Kriegman, "Two faces are better than one: Face recognition in group photographs," in *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–8, IEEE, 2011.
- [15] Y.-H. Lei, Y.-Y. Chen, L. Iida, B.-C. Chen, H.-H. Su, and W. H. Hsu, "Photo search by face positions and facial attributes on touch devices," in *Proceedings of the 19th ACM international conference on Multimedia*, pp. 651–654, ACM, 2011.
- [16] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, "Robust gender recognition by exploiting facial attributes dependencies," *Pattern Recognition Letters*, vol. 36, pp. 228–234, 2014.
- [17] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015.
- [18] Y. Zhong, J. Sullivan, and H. Li, "Leveraging mid-level deep representations for predicting face attributes in the wild," in *Image Processing (ICIP), 2016 IEEE International Conference on*, pp. 3239–3243, IEEE, 2016.
- [19] E. M. Rudd, M. Günther, and T. E. Boulton, "Moon: A mixed objective optimization network for the recognition of facial attributes," in *European Conference on Computer Vision*, pp. 19–35, Springer, 2016.
- [20] M. A. Haque, K. Nasrollahi, and T. 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*, pp. 443–448, IEEE, 2013.
- [21] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, vol. 3, no. 4, pp. 314–324, 2014.
- [22] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. 1–1, IEEE, 2002.
- [23] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2879–2886, IEEE, 2012.
- [24] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. 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*, pp. 1–8, IEEE, 2013.
- [25] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. 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.