2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Learning a Lightweight Deep Convolutional Network for Joint Age and Gender Recognition

Linnan Zhu*, Keze Wang*†, Liang Lin† and Lei Zhang*
*Department of Computing, The Hong Kong Polytechnic University, China
Email: {cslzhu, cskwang, cslzhang}@comp.polyu.edu.hk
†Sun Yat-sen University, China
Email: linliang@ieee.org

Abstract—**This paper proposes a lightweight deep model to recognize age and gender from a face image. Though simple, our network architecture is able to complete the two tasks effectively and efficiently. Moreover, different from existing methods, we simultaneously perform the age and gender recognition tasks via a joint regression model. Specifically, our model employs a multi-task learning scheme to learn shared features for these two correlated tasks in an end-to-end manner. Extensive experimental results on the recent Adience benchmark demonstrate that our model achieves competitive recognition accuracy with the state-of-the-art methods but with much faster speed, i.e., about 10 times faster in the testing phase. Our model can be easily adopted and extended to other facial applications.**

1. Introduction

Real time facial attribute recognition is a very promising and hot topic, especially, recognizing age and gender in a single image has sparked off a great interest in both research community and industrial companies. For the age recognition task, researchers firstly employ the hand-craft local features for representing the distribution of face images, such as Gaussian Mixture Models (GMM) [1], and Hidden-Markov-Model [2]. After that, they further presented to use different feature descriptors, for the purpose of representing the face image, e.g. Gabor feature [3], Biologically-Inspired features (BIF) [4], local binary patterns (LBP) [5], followed by Support Vector Machines. For the task of gender recognition, the research road map is very similar to the development of age recognition, because they are highly correlated tasks and both of them belong to the set of facial attributes. [6] used image intensities followed by SVM classifiers and [7] implemented Webers' Local texture Descriptor [8], both of which are based on human hand-craft features.

More recently, the popular deep learning technique achieves incredible progress in visual recognition [9], and also has been successfully applied to age and gender recognition. Levi et al. [10] proposed to individually train two models for each problem, which can be treated as a cascade of Convolutional Neural Networks (CNN). However, these CNN based methods are time consuming for mobiles or low-end PCs for the following two issues:

- Exploiting the complex CNN architecture. Recently, many CNN based methods directly implement the popular architectures (AlexNet [9] and VGG [11]), which are specially designed for large scale visual recognition, e.g. ImageNet Challenge [12]. For the case of age and gender recognition tasks, however, these network architectures are too complex and over-designed. This heavily increases the computation burden.
- Regarding age and gender recognition as two independent problems. Although in [10] the model they trained for age and gender recognition has a same architecture, the parameters are different and the method requires a complex cascade architecture of deep model. As the matter of fact, age and gender recognition are two highly correlated tasks about facial attributes. It will be beneficial to recognize accuracy and time efficiency if we can optimize these two tasks together.

To address above mentioned issues, we propose a lightweight deep framework to jointly recognize age and gender in a fast end-to-end manner. The proposed framework employs a multi-task learning scheme to complete these two correlated tasks. As is known to all, the effectiveness of multi-task learning has been verified on many computer vision problems, e.g. image classification [13], visual tracking [14] and facial landmark detection [15]. On these problems, multi-task learning achieves better performance than training a single task one by one. The reason is that multi-task learning can exert a conducive position on extracting the shared feature to improve the accuracy of each task, which can be better than optimize each task. Moreover, it is very easy to extend our model for other more tasks, e.g. facial expression and other attributes.

The **key contributions** of this paper are listed as follows: 1) To our knowledge, this is the first attempt to investigate how age and gender recognition can be optimized together to learn a correlated multi-task. Our multi-task learning scheme enables to share and learn optimal features to improve recognition performance for both two tasks. Notably, the proposed model does not limit the number of related tasks, we can extend to many other

978-1-5090-4847-2/16/$31.00 ©2016 IEEE

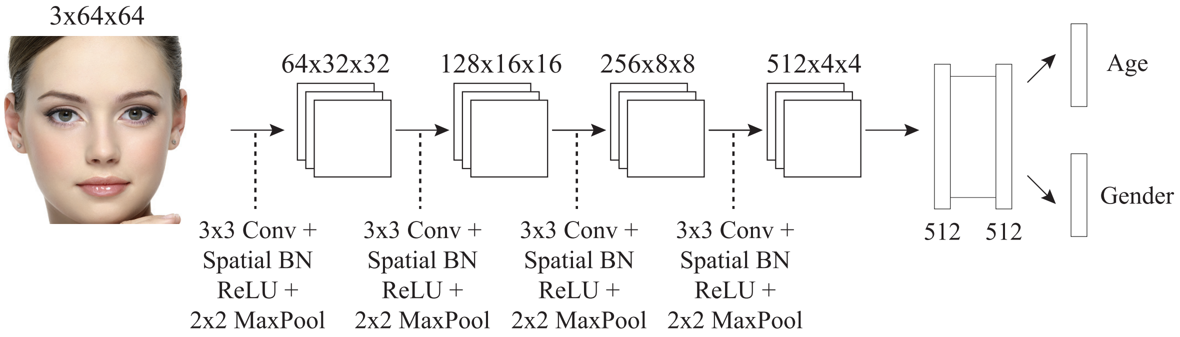

Figure 1. The architecture of our lightweight deep model for the age and gender recognition. The network consists of four convolution operations and two fully connected layers, where the raw image pixel are treated as the input.

tasks, e.g. facial expression and other attributes. 2) The network architecture employed by our model is specially designed for age and gender recognition to improve the time efficiency while keeping the quality of recognition performance. Not only outperforming the compared methods [10], [16] in recogntion accuracy, the experimtental results but also demonstrate that our model runs 9 times faster than [10] and achieves real-time performs even on a low-end PC or a mobile device. Thus, it is very suitable for our model to be implemented in the commercial applications or industry.

The remainder of the paper is organized as follows. Sect. 2 presents a review of related works. Then we present our lightweight deep convolutional network in Sect. 3, including the network architecture, multi-task learning scheme, and the training/testing procedure of our model. The experimental results, comparisons and component analysis are exhibited in Sect. 4. At last, Sect. 5 concludes this paper.

2. Related Work

2.1. Age Recognition

For the age recognition task, Kwon et al. [17] firstly published papers in this filed. They applied cranio-facial development theory to distinguish baby and adults by calculating six ratios of distance on frontal face images. Then, Ramanathan et al. [18] attempted eight ratios of distance. Nevertheless, due to the anthropometry concept based models are very sensitive to head pose and do not use the additional texture information, this model is not suitable to distinguish the adults. [19] proposed an active appearance model (AAM) to represent the face image. Lanitis et al. [20] tried different classifiers for age estimation based on their age image representation, especially the quadratic aging function [21]. The advantage of the AAM model over the previous anthropometry based model is that the AMM model used shape and texture information together and can classify different age groups, not only the young and the adults. Then Geng et al. [22] proposed an AGing pattErn Subspace (AGES) model to learn a personalized pattern for a individual person. However, this method needs a large number of cross-age images of one person. [23] proposed to learn a low dimensional pattern from face images at each age, which makes it much easier and more flexible to build a face database. Afterwards, some feature extraction related algorithms were introduced to this area. The local features, such as texture and shape features [24], and Local Binary Patterns (LBP) [5] were used first, but due to the obvious local variations in facial appearance, some global feature were applied afterwards. Spatially Flexible Patch (SFP) feature put forward by Yan et al. [25] includes the local patches and position information, aiming to help to deal with occlusion, and head pose variations. In the previous study, researchers calculated ratios between different facial features [26]. After that, [22], [27] presented subspace and manifold learning respectively. They obtained good performance on some certain constrained datasets, i.e., the face images are frontal and well-aligned, which are not suitable for the unconstrained face benchmark. The above mentioned approaches are all evaluated on some certain benchmarks, which have a lot of constrains, such as face images need to be captured in a near-front view and with a suitable lighting, well-aligned or other perfect conditions. In a word, these related methods can not be directly implemented for the real-world challenge.

2.2. Gender Recognition

Along the development of age recognition, the research on gender recognition is back to 1990s. Cottrell et al. [28] first proposed a neural network model, but the faces were under constrained. Then Brunelli et al. [29] proposed a HyperBF network for this task by extracting the geometrical features. In [30], Wiskott et al. implemented Gabor wavelets to a face representation. Afterwards, [31] applied principal component analysis (PCA) and linear discriminant analysis (LDA) based on Gabor wavelet to recognize the gender label. Then Sun et al. [32] demonstrated that genetic algorithms (GA) was suitable for gender recognition task and verified the feature selection was at a very important stage for this task. And In [33], they obtained good recognition accuracy in a constrained dataset FERET. This method extracts feature via independent component analysis and uses LDA to be the classifier. Afterwards, Costen et al. [34] proposed a sparse SVM approach and yielded good results in a Japanese face dataset. In [35] researchers implemented SVM and Adaboot classifiers on the raw images, respectively. In the recent years, Webers Local texture Descriptor [8] was introduced in gender recognition task and yielded good performance on a constrained face images dataset.

2.3. Deep Convolutional Neural Networks

By directly extracting features from raw images, deep CNN models have made impressive progresses on visual recognition problems, e.g, image classification, object detection, semantic segmentation and many other recognition [9], [11]. Inspired by the success of CNNs on visual recognition problems, Levi et al. [10] proposed a deep CNN model to tackle the age and gender recognition problem and obtained significant results by sufficient training data. Because of the high resolution of input image (227×227) and large convolution filters , the parameters of this network architecture is in a large quantity. Besides, the CNN based approach are complex enough and hardly to implement on a low-end PC or a mobile phone. In order to deal with this problem, in this paper, we propose a lightweight deep model for the age and gender recognition tasks.

3. Lightweight Deep Convolutional Network

In this section, we will present our proposed lightweight deep convolution neural networks for Age and Gender recognition. The framework of our model consists of two stages: shared feature extraction and multi-task estimation stage. In the following, we will describe our model from two aspects: Network architecture and multi-task learning scheme.

3.1. Network Architecture

As illustrated in Fig. 1, our proposed model is constructed by stacking four spatial convolution operations and two fully connected layers. The convolution operation in this work includes Convolution (Conv) + Spatial Batch Normalization (Spatial BN) + Rectified Linear Unit (ReLU) + Max Pooling (MaxPool). The number of convolution filters for the four convolution operations are 64 with $3 \times 3 \times 3$ size, 128 with $64 \times 3 \times 3$ size, 256 with $128 \times 3 \times 3$ size, 512 with $256 \times 3 \times 3$ size, respectively. The two fully connected layers both have 512 neurons. The input RGB image is downsampled into the size 64×64 before being fed into the network. Notably, our experiments demonstrate that 64×64 resolution is good enough for age and gender recognition tasks. Then the network regresses 2-dimension vector with the estimated age and gender labels for the input image. In summary, our network architecture supports low resolution of the input image and consists of small convolutional filters and fully connectedly layers. This implies that our proposed network is lightweight.

Thanks to the lightweight design, our network architecture has only 6×10^6 parameters and is significantly lightweight, compared with the popular architecture AlexNet [9] (60×10^6 parameters, $10\times$ bigger than us) and VGG-16 [11] (138×10^6 parameters, $23\times$ bigger than us). That is to say, our proposed method is simple but effective, and have great impact on solving the time consuming problem. Hence, our network is able to be implemented in a low-end PC or even a mobile device.

3.2. Multi-task Learning Scheme

In order to jointly perform age and gender recognition, we exploit the multi-task learning scheme by regarding these two correlated tasks as a regression problem. Specifically, built upon the last fully connected layer, the output of our model is a regressed label vector with two prediction results for age and gender, respectively. In this way, both the age and gender tasks share the same feature representation. The goal of our network is to learn the shared feature to complete these two correlated tasks. Many research [13], [14], [15] clarify that the multi-task learning scheme is able to improve the generalization performance of multiple related tasks. In the following section, we will describe the formulation of our exploited mult-task learning scheme.

Suppose we have N training samples, C tasks ($C = 2$ in this work) to be completed and y_i^c is the ground truth label of the c-th task for the i-th image I_i. Thus the objective of our proposed multi-task learning scheme is defined as:

$$\arg\min_{\{\omega, \mathbf{w}_c\}} \sum_{i=1}^{N} \sum_{c=1}^{C} l\left(y_i^c, \phi(I_i, \omega); \mathbf{w}_c\right) + \Psi(\mathbf{w}_c), \quad (1)$$

where $\phi(I_i, \omega)$ denotes the feature vector of our model, ω is the corresponding network parameter. $\mathbf{w}_c$ is the regression parameter for the c-th task. The $\Psi(\mathbf{w}_c)$ is the L_2 norm regularization term that penalizes the complexity of $\mathbf{w}_c$ to avoid model overfitting, i.e., $\Psi(\mathbf{w}_c) = \|\mathbf{w}_c\|_2^2$. The function $l(\cdot, \cdot)$ denotes the estimation error for label regression, and is defined as follows:

$$l\left(y_i^c, \phi(I_i, \omega); \mathbf{w}_c\right) = \|y_i^c - \mathbf{w}_c^T \phi(I_i, \omega)\|_2^2. \quad (2)$$

To this end, we have presented our deep model for age and gender recognition. Our model has a lightweight architecture and jointly optimizes these two tasks by learning a shared feature representation for regression. In the next subsection, we will discuss the training and testing phase of our proposed model.

3.3. Model Training and Testing

As our proposed model regards the age and gender recognition task as a regression formulation, the standard back propagation algorithm [36] is applicable to optimize

Figure 2. Example face images for age and gender recognition from the Adience benchmark.

the model parameters $\{\omega, \{\mathbf{w}_c\}_{c=1}^{C}\}$. Specifically, the partial derivatives with respect to $\{\omega, \{\mathbf{w}_c\}_{c=1}^{C}\}$ are defined as:

$$\begin{aligned} \partial \frac{l\left(y_i^c, \phi(I_i, \omega); \mathbf{w}_c\right)}{\omega} &= 2\left(y_i^c - \mathbf{w}_c^T \phi(I_i, \omega)\right) \partial \frac{\phi(I_i, \omega)}{\omega} \\ \partial \frac{l\left(y_i^c, \phi(I_i, \omega); \mathbf{w}_c\right)}{\mathbf{w}_c} &= -2\left(y_i^c - \mathbf{w}_c^T \phi(I_i, \omega)\right) \phi(I_i, \omega). \end{aligned} \tag{3}$$

Once all of above mentioned derivatives are obtained, we can perform the stochastic gradient descending method to update the model parameters $\{\omega, \{\mathbf{w}_c\}_{c=1}^{C}\}$.

In the testing phase, given an input image, our model can directly output both the age and gender estimation result through forwarding the network.

4. Experiments

4.1. Dataset Description and Setting

To verify the effectiveness and efficiency of our proposed model, we conduct experiments on the recently released Adience benchmark [37], which is mainly constructed for age and gender recognition and contains 26K images of 2,284 different people with the resolution 816×816. The Adience benchmark is very challenge because its images are directly collected from mobile devices. As illustrated in Fig. 2, these images are highly unconstrained with extreme variations in head pose, lightning conditions quality, which can represent the real-world challenge. There are 8 categories (i.e., *0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-*) to represent the age level of the subjects in Adience benchmark.

Our model is trained from scratch and no outside data are used in training phase. Stochastic gradient decent (SGD) is employed to optimize the parameters with image batch size of 128 images. The initial learning rate is 0.1, reduced to one tenth after 25 iterations. The momentum is 0.9. The weight decay is 5×10^{-4}. The learning rate decay is 10^{-7}. Dropping out strategy is also used in the fully connected layers with 0.5 ratio. Training the network on the Adience benchmark takes around 1 hour.

We compare our model with three state-of-the-art methods [7], [10], [16] from both accuracy and efficiency. Adopting the evaluation protocol of [10], we perform a standard five-fold, subject-exclusive cross-validation to obtain the recognition accuracy of age and gender. Recognition accuracy is defined as the number of corrected classified samples divides total sample number.

TABLE 1. COMPARISON OF AVERAGE GENDER AND AGE RECOGNITION RESULTS. THE ENTRIES WITH BEST ACCURACY ARE BOLD-FACED.

Method	Gender	Age
Sun et al. [16]	77.8 ± 1.3	45.1 ± 2.6
Ullah et al. [7]	79.3 ± 0.0	-
Levi et al. [10]	85.9 ± 1.4	**49.5** ± 4.4
Ours	**86.0** ± 1.2	46.0 ± 1.1

TABLE 2. COMPARISON OF THE AVERAGE RUNNING TIME (SECOND PER IMAGE) FOR DIFFERENT NETWORK ARCHITECTURE.

Architecture	AlexNet [9]	VGG [11]	Levi et al. [10]	Ours
PC	0.15	0.45	0.09	0.01
Mobile	inapplicable	inapplicable	5.2	0.5

4.2. Comparison Results

Recognition Accuracy: Tab. 1 illustrates the recognition accuracy on the Adience benchmark. As is reported in Table. 1, the prediction accuracy of [7], [10], [16] and our model for gender recognition are 77.8%, 79.3%, 85.9% and 86.0%, respectively. This result demonstrates that our model can obtain the comparable performance with the state-of-the-art methods. Some correctly predicted samples and false estimations are illustrates in Fig. 3.

Time efficiency: Firstly, we compare the running time of ours with several CNN architectures, e.g., AlexNet [9], VGG [11], [10], for age and gender estimation on PC platform, which is a desktop with intel 4.0 Ghz CPU and nvidia 970 GPU. Given an input image, the average running times are illustrated in the first row of Tab. 2. It is obvious that our model is $15\times, 45\times, 9\times$ faster than the compared AlexNet [9], VGG [11], [10] on the PC platform, respectively.

Meanwhile, considering age and gender recognition are very promising applications in mobile platform, we choose the fastest two models, i.e., [10] and ours to implement on a Samsung Note 3 mobile phone, which has very limited computation capability. As for a mobile application, the processing time is a key factor for users. We can see from the second row of Tab. 2, though obtaining slightly better performance, the compared method [10] requires about 5 seconds for a single image. This processing time is not acceptable. Thanks to the lightweight advantage, Tab. 2 demonstrates that our model only costs 0.5s to predict age and gender for a single image, and is about 10 times faster than [10]. The speed advance is

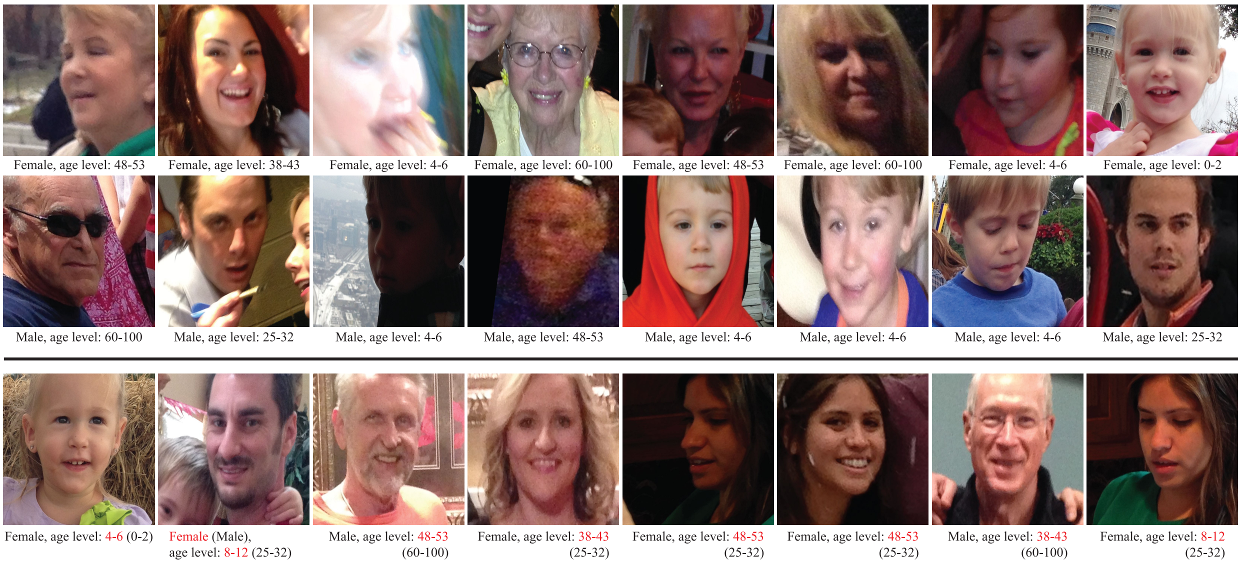

Figure 3. The age and gender estimations of our proposed model. The samples in the first and second rows demonstrate that the age and gender are correctly predicted in black, while the last row shows failure cases with wrong prediction in red.

TABLE 3. COMPONENT ANALYSIS OF SINGLE AND MULTIPLE TASK WITH DIFFERENT NUMBER OF CONVOLUTION OPERATIONS.

Method	Gender	Age
Ours (single-6conv)	85.3 ± 0.8	$\mathbf{49.7} \pm 0.6$
Ours (single-5conv)	85.0 ± 1.0	49.0 ± 0.5
Ours (single-4conv)	84.4 ± 0.5	48.3 ± 0.7
Ours (single-3conv)	83.7 ± 0.7	47.6 ± 0.8
Ours	$\mathbf{86.0} \pm 1.2$	46.0 ± 1.1

due to that our proposed model has only a few parameters and supports low-solution images. Hence, our model significantly outperforms [10] on the running time.

From the aspect of effectiveness and efficiency, the experimental results validate the contribution of our network architecture. Thanks to the lightweight design, our model is able to be applied to common mobile devices.

4.3. Component Analysis

To specify the contribution of our employed multi-task scheme, we have conducted the following experiments. We construct the single task version of our model, i.e., we discard the multi-task scheme and separately train a single model for gender and age recognition. Moreover, we also investigate that the performance of convolution layer number from 3 to 6. We denote them as Ours ("single-3conv", "single-4conv", "single-5conv" and "single-6conv"), respectively. Notably, Ours "single-4conv" has the same architecture of our multi-task model, while "single-5conv" and "single-6conv" has one and two more convolution + ReLU (filters are 512 with $512 \times 3 \times 3$ size). The experiment results are illustrated in Tab. 3, where we can obtain two observations: i) As the network becomes deeper, the gender and age recognition accuracy increases from 83.7% to 85.3% and 47.6% to 49.7%, respectively. This meets the "the deeper, the better" conclusion; ii) With only 4 convolution layers, our model achieves the best performance on the gender recognition. This justifies the effectiveness of the employed multi-task scheme for the gender recognition. The multi-task scheme can simplify the CNN architecture and improve the performance together.

5. Conclusions

In this paper, we investigated how age and gender recognition can be optimized jointly via a lightweight deep model. Not only obtaining competitive performance with the state-of-the-art methods, our proposed approach also runs much faster. Moreover, thanks to the proposed multi-task learning scheme, our model can be easily extended to other facial attribute recognition tasks, e.g., facial expression, face recognition and facial similarity.

Acknowledgement

This work is supported by the Hong Kong RGC GRF grant (PolyU 152124/15E).

References

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition (2Nd Ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.

[2] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," *IEEE ASSp Magazine*, 1986.

[3] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, Apr 2002.

[4] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, pp. 1019–1025, 1999.

[5] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec 2006.

[6] B. Moghaddam and M.-H. Yang, “Learning gender with support faces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, May 2002.

[7] I. Ullah, M. Hussain, G. Muhammad, H. Aboalsamh, G. Bebis, and A. M. Mirza, “Gender recognition from face images with local wld descriptor,” in *2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, April 2012, pp. 417–420.

[8] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, “Wld: A robust local image descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, Sept 2010.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[10] G. Levi and T. Hassncer, “Age and gender classification using convolutional neural networks,” in *2015 IEEE Conference o [13]n Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 34–42.

[11] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.

[12] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[13] X. T. Yuan and S. Yan, “Visual classification with multi-task joint sparse representation,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 3493–3500.

[14] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Robust visual tracking via structured multi-task sparse learning,” *International Journal of Computer Vision*, vol. 101, no. 2, pp. 367–383, 2012.

[15] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI*. Cham: Springer International Publishing, 2014, ch. Facial Landmark Detection by Deep Multi-task Learning, pp. 94–108.

[16] Y. Sun, X. Wang, and X. Tang, “Deep convolutional network cascade for facial point detection,” in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3476–3483.

[17] Y. H. Kwon and N. da Vitoria Lobo, “Age classification from facial images,” in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR ’94., 1994 IEEE Computer Society Conference on*, Jun 1994, pp. 762–767.

[18] N. Ramanathan and R. Chellappa, “Modeling age progression in young faces,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 1, June 2006, pp. 387–394.

[19] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, Jun 2001.

[20] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing different classifiers for automatic age estimation,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 621–628, 2004.

[21] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Toward automatic simulation of aging effects on face images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, Apr 2002.

[22] X. Geng, Z. H. Zhou, and K. Smith-Miles, “Automatic age estimation based on facial aging patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, Dec 2007.

[23] Y. Fu, Y. Xu, and T. S. Huang, “Estimating human age by manifold analysis of face pictures and regression on aging features,” in *2007 IEEE International Conference on Multimedia and Expo*, July 2007, pp. 1383–1386.

[24] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu, “Method for estimating and modeling age and gender using facial image processing,” in *Virtual Systems and Multimedia, 2001. Proceedings. Seventh International Conference on*, 2001, pp. 439–448.

[25] S. Yan, M. Liu, and T. S. Huang, “Extracting age information from local spatially flexible patches,” in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 737–740.

[26] Y. H. Kwon and N. da Vitoria Lobo, “Age classification from facial images,” in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR ’94., 1994 IEEE Computer Society Conference on*, Jun 1994, pp. 762–767.

[27] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, “Image-based human age estimation by manifold learning and locally adjusted robust regression,” *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, July 2008.

[28] G. W. Cottrell and J. Metcalfe, “Empath: Face, emotion, and gender recognition using holons,” in *Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems 3*, ser. NIPS-3, 1990, pp. 564–571.

[29] B. Poggio, R. Brunelli, and T. Poggio, “Hyberbf networks for gender classification,” in *Proc. DARPA Image Understanding Workshop*, 1995.

[30] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” in *Image Processing, 1997. Proceedings., International Conference on*, vol. 1, Oct 1997, pp. 129–132 vol.1.

[31] M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, “Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 202–207.

[32] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, “Genetic feature subset selection for gender classification: a comparison study,” in *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*, 2002, pp. 165–170.

[33] A. Jain and J. Huang, “Integrating independent components and linear discriminant analysis for gender classification,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, May 2004, pp. 159–163.

[34] N. P. Costen, M. Brown, and S. Akamatsu, “Sparse models for gender classification,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, May 2004, pp. 201–206.

[35] B. Moghaddam and M.-H. Yang, “Learning gender with support faces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, May 2002.

[36] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, “Handwritten digit recognition with a back-propagation network,” in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed., 1990, pp. 396–404.

[37] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of unfiltered faces,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, Dec 2014.