Facial Age Estimation With Age Difference

Zhenzhen Hu, Yonggang Wen, *Senior Member, IEEE*, Jianfeng Wang, Meng Wang, *Member, IEEE*, Richang Hong, *Member, IEEE*, and Shuicheng Yan, *Fellow, IEEE*

***Abstract*— Age estimation based on the human face remains a significant problem in computer vision and pattern recognition. In order to estimate an accurate age or age group of a facial image, most of the existing algorithms require a huge face data set attached with age labels. This imposes a constraint on the utilization of the immensely unlabeled or weakly labeled training data, e.g., the huge amount of human photos in the social networks. These images may provide no age label, but it is easy to derive the age difference for an image pair of the same person. To improve the age estimation accuracy, we propose a novel learning scheme to take advantage of these weakly labeled data through the deep convolutional neural networks. For each image pair, Kullback–Leibler divergence is employed to embed the age difference information. The entropy loss and the cross entropy loss are adaptively applied on each image to make the distribution exhibit a single peak value. The combination of these losses is designed to drive the neural network to understand the age gradually from only the age difference information. We also contribute a data set, including more than 100 000 face images attached with their taken dates. Each image is both labeled with the timestamp and people identity. Experimental results on two aging face databases show the advantages of the proposed age difference learning system, and the state-of-the-art performance is gained.**

***Index Terms*— Age estimation, age difference, convolutional neural networks, K-L divergence distance.**

I. Introduction

AS AN important biological information carrier, the human face reflects lots of properties such as identity, age, gender, expression, and emotion. With the passage of time, the facial appearance changes as human aging, which indicates human behaviour and preference. Although different people are aging differently and aging shows various forms in different ages, there are still some general changes and resemblances we can always describe [1]. Human age can be directly inferred by distinct patterns from the facial appearance. For the same person, the photos taken at different years reveal the aging process on their faces. The longer the interval is, the more obvious changes there will be, as shown in Figure 1. Age information plays an important role in human-computer interaction and Artificial Intelligence systems and shares many in other face-related tasks such as face detection and recognition. Image based human age estimation has wide potential practical applications, e.g., demographic data collection for supermarkets or other public areas, age-specific human computer interfaces, age-oriented commercial advertisement, and human identification based on old ID-photos.

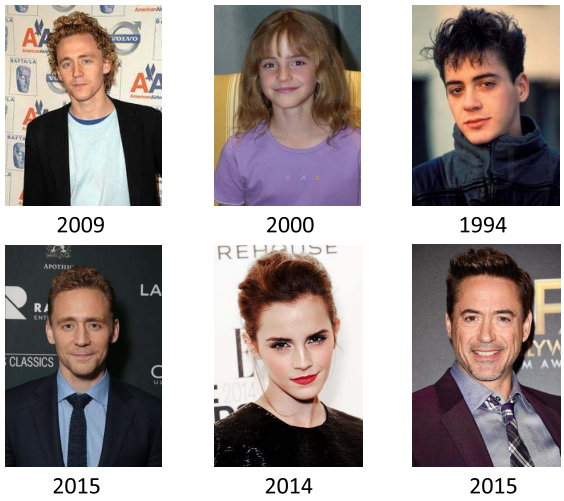


Fig. 1. Photos from the same person of different years reveal the changing process of aging. Each column shows the images of the same person and below the photo is the taken year.

Estimating age from images has been historically one of the most challenging problems within the field of facial analysis. With the rapid advances in computer vision and pattern recognition, computer-based age estimation on faces becomes a particularly interesting topic. However, human estimation of facial age is usually not as accurate as other kinds of facial information such as identity and gender. It is very challenging to accurately predict the age of a given facial image because human facial aging is generally a slow and complicated process influenced by many internal and external factors.

In recent years, the interest in human facial age estimation has significantly increased [2]–[6]. A typical pipeline of the existing methods for age estimation usually consists of two modules [7]: (1) extracting image features/representations for age, and (2) learning an age estimator with these image features. Various facial age features have been developed

Manuscript received May 20, 2016; revised September 23, 2016; accepted November 17, 2016. Date of publication December 1, 2016; date of current version May 9, 2017. This work was supported by Nanyang Technological University under Grant AcRF Tier 1 RG26/16 and Grant Tier 2 ARC 42/13. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Weisheng Dong.

Z. Hu and Y. Wen are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: huzhen.ice@gmail.com; ygwen@ntu.edu.sg).

J. Wang is is with Azure Storage, Microsoft, Seattle, WA 98502 USA (e-mail: wjf2006@mail.ustc.edu.cn).

M. Wang and R. Hong are with the Hefei University of Technology, Hefei 230009, China (e-mail: eric.mengwang@gmail.com; hongrc@hfut.edu.cn).

S. Yan is with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: eleyans@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2633868

Fig. 2. The processing of human face aging. These examples are aging faces of one subject in the FG-NET database.

Fig. 3. These examples are aging faces of four subjects in the MOPRPH database. Each pair is from the same person.

for facial age estimation. Among them, biologically inspired features (BIF) proposed by Guo *et al.* [8] shows the best performance on age estimation and has been wildly used. With the obtained image features for age, various methods have been proposed to learn an age estimator. In most of these methods, age estimation is regarded as either a classification problem [9] or a regression problem [10]. Recently, deep learning schemes, especially Convolutional Neural Networks (CNNs), have been successfully employed for many tasks related to facial analysis, including face detection, face alignment [11], face verification [12], and demographic estimation [13]. Wang *et al.* [14] extracted feature maps obtained in different layers as age features based on the deep learning model. Huerta *et al.* [4] provided a thorough evaluation on deep learning for age estimation and compared it with the hand-crafted fusion features.

Although a number of algorithms have been successfully developed for facial age estimation, many challenges still remain. First of all is the insufficient labeled data to cover all age patterns. The purpose of age estimation is to automatically label a facial image with exact age (year) or age group (year range) [7]. The patterns of human age are complex, and it is difficult for supervised methods to cover aging patterns. Especially with the popularization of deep neural networks in computer vision, a large-scale labeled dataset becomes more significant. The commonly-used datasets for age estimation, e.g., FG-NET Aging Database [15] and MORPH [16], contain 50,000 face images mainly frontal-view and neutral, which are much less than the conventional datasets such as ImageNet. In FG-NET dataset, each subject has a lot of samples across youth to old age, as shown in Figure 2. For MORPH dataset, as shown in Figure 3, a lot of subjects only have two or three images. However, a large aging database is hard to collect, especially the chronometrical image series for an individual. Besides, manual labels are costly and sometimes impractical to obtain after system deployment. To address this issue, some research works start to take advantage of the information from

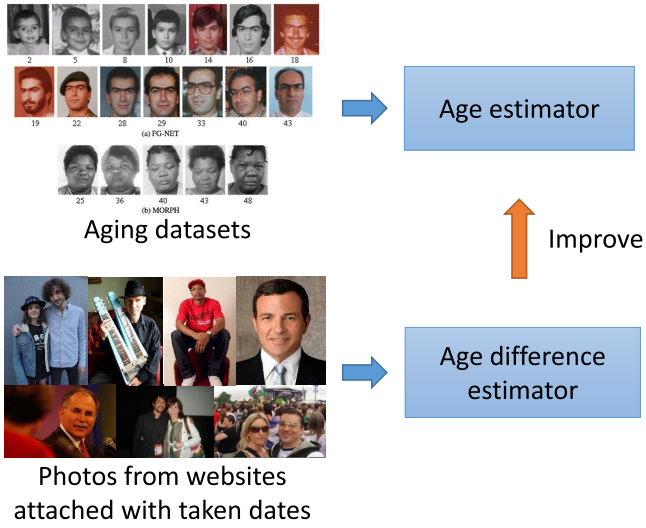


Fig. 4. Schematic illustration of estimating human age through the image without age label.

other sources such as videos [13]. Chen *et al.* [17] utilized cumulative attribute space to solve the noisy and sparse low level aging features regression. Li *et al.* [18] proposed a semi-supervised learning mode to learn the underlying structural information of unlabeled data. Fortunately, although it is hard to obtain sufficient age-labeled data, there are numerous accessible human face photos on the Internet. People share their photos on social networks and most of these photos are attached with the uploading date. The different photos taken in different years can reveal the age information of people.

In this paper, we investigate the problem of age estimation without sufficient ground truth labels and propose an approach to estimate the age of a human face with the assistance of age difference information. The brief illustration is shown in Figure 4. With the explosion of social networks, there are massive human photos uploaded immediately after taken. We contribute a face dataset including more than 150,000 images of four thousand subjects from the photo share website Flickr attached with their taken date. These images are downloaded automatically via query words of the subjects name from LFW dataset [19]. Each image is labeled with both taken year and people identity. All the images are filtered with face detection and alignment algorithms to guarantee the reliability of the dataset. Given a pair of face images taken at different years from the same subject, we explore the age difference of images via the deep Convolutional Neural Networks. First, we build a deep age estimator based on the standard aging datasets. A symmetric Kullback-Leibler divergence loss function is placed at the top layer of CNNs. We utilize label distribution to design the

loss function. For the non age-labeled image, we combine the images of the same subject into pairs and take the difference of taken years as the difference of ages. The pairs of images are used to fine tune the pre-trained deep age estimator. In this step, we design three kinds of loss functions on the top of the softmax layer, i.e. entropy loss, cross entropy loss and K-L divergence distance, to learn the representation of the age difference. The combination of loss functions can force the probability distribution of age classes to have a single peak value at the location of the correct range. The deep CNN age estimator is first trained on the standard aging dataset with age labels and then improved by the non-age-label dataset.

The main novelties and contributions of this work are three-fold:

1) We propose an approach to exploit human age information through the age difference. To our best knowledge, it is the first time that the age difference information has been used for humane age estimation.
2) We explore the age difference information with three kinds of loss functions, i.e. entropy loss, cross entropy loss and K-L divergence distance. These loss functions can not only force the probability distribution of age classes to have one single peak value but also make the probability distribution locate within the correct range. The proposed approach will be an important component in practical systems for age estimation, which is able to address the human faces with arbitrary poses, arbitrary age, and arbitrary ethnicity.
3) We also contribute a dataset containing more than 100,000 face images attached with their taken dates. Each image is labeled with both the timestamp and people identity.

The rest of this paper is organized as follows. In Section II, we give a brief review of related work on age estimation. Then we provide a detailed description of our approach and show how to utilize the age difference information to improve the age estimation in Section III. The experiments are reported in Section IV. Finally, we draw the conclusion of this work in Section V.

II. Related Work

In the past few years, much research has been conducted in human facial age estimation. The earliest paper published in the area of age classification from facial images was the work by Kwon and Lobo [20]. They proposed a human age classification method based on the cranio-facial development theory and skin wrinkle analysis, where the human faces are classified into three groups, namely, babies, young and senior adults. Lanitis *et al.*[21] and [22] adopted the statistical face model, Active Appearance Models (AAMs) [23], to extract the shape and texture information of facial images. In their work, the aging pattern is represented by a quadratic function called the aging function. Later, Geng *et al.* [24] and [25] proposed the AGing pattErn Subspace (AGES) algorithm based on the subspace trained on a data structure called aging pattern vector. Yan *et al.* [10] regarded age estimation as a regression problem with nonnegative label intervals and solved the problem through semidefinite programming. They also proposed an EM algorithm to solve the regression problem and speed up the optimization process [26]. Instead of learning a specific aging pattern for each individual, a common aging trend or pattern can be learned from many individuals at different ages. One possible way to learn the common aging pattern is the age manifold [27], which utilizes a manifold embedding technique to learn the low-dimensional aging trend from many face images at each age. After that, various aging features were developed for facial age estimation. Aging-related facial feature extraction is more focused by the appearance model. Hayashi *et al.* [28] considered both texture (wrinkle) and shape (geometry) features to characterize each face image. The effective texture descriptor, Local Binary Patterns (LBP), has been used for appearance feature extraction in an automatic age estimation system [29]. The Gabor feature has also been tried on the age estimation task [30], which has been demonstrated to be more effective than LBP. The biologically inspired features were introduced to the age estimation by Guo *et al.* [8], where age features were extracted with a set of predefined Gabor filters with different scales and orientations.

To build a robust facial age estimation system, Ni *et al.* [31] proposed a method based on the mining of the noisy aging face images collected from the web images and videos. Chang *et al.* [32] transformed an age estimation task into multiple cost-sensitive binary classification subproblems, and solved the problem with an ordinal hyperplane ranking algorithm. One of the most recent progresses was made by Geng *et al.* [2], who introduced the label distribution for the age estimation problem. They regarded each face image associated with a label distribution. That is, each image can contribute to learning its chronological age and adjacent ages. They proposed two algorithms, named IIS-LLD and CPNN, for solving the multi-label problem. In this paper, we also utilize the label distribution algorithm for the labeled images.

Deep learning methods, which can automatically learn the effective image representations, have achieved a great success in object classification [33]. Recently, deep learning schemes, especially CNNs, have been successfully employed for many tasks related to facial analysis, including face detection, face alignment [11], face verification [12], and demographic estimation [13]. Wang *et al.* [14] extracted feature maps obtained in different layers as age features based on the deep learning model. Huerta *et al.* [4] provided a thorough evaluation on deep learning for age estimation and compared it with the hand-crafted fusion features. The recent trend is that high-quality learned features will replace the hand-crafted features. The fixed hand-crafted visual features may not be optimally compatible with the aging process. Ideally, it is expected that age image representation can sufficiently preserve the age information. Current deep learning tends to increase the depth of networks. GoogLeNet [34] is a new Inception-style architecture which is used to increase the representational power of networks. Inspired by this, we use deep learning to simultaneously learn the image representation and age estimation. We demonstrate that switching from the hand-crafted image features to the deep learned features together with improvements in the objective functions and a huge

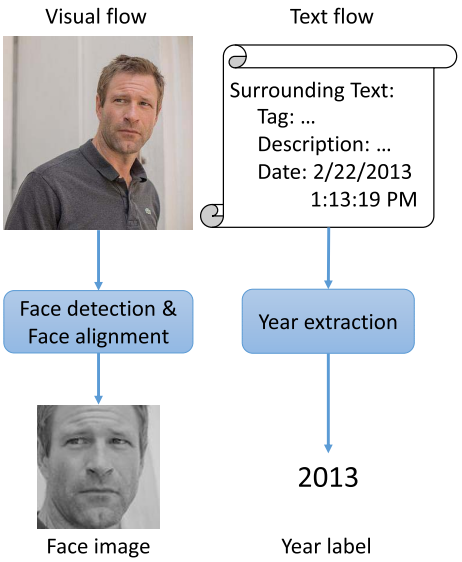


Fig. 5. The age difference dataset construction.

amount of unlabeled video information leads us to state-of-the-art age estimation performance.

Another similar research topic in recent years is apparent age estimation. In the competition organized by ChaLearn [35], age is labeled by different volunteers given only the images containing the single individuals. Compared with real age, the annotated apparent age could be mutable, but the mean of labels from different annotators are highly stable and thus can be defined as the apparent age. Liu *et al.* [36] and Yang *et al.* [37] both adopted the deep learning framework and label distribution learning for apparent age estimation. In the work of [36], they combined age classifier and age regression models based on GoogLeNet. For the age regressor, the Euclidean loss was used to measure the 1-dimensional real-value encoding. For age classifier, they replaced the loss layer with cross-entropy loss. Different from these apparent age estimation research, our paper focus on the real age estimation problem. Although these two problems may be similar with each other, there are still some differences. First, there is no ground truth age of apparent age dataset and the age annotation is very subjective. Second, in the challenge of ChaLearn, the measurement of apparent age estimation is mean normalized error, while for the real age estimation is mean absolute error. The age estimator proposed by [36], [37] is not suit for standard aging dataset, such as FGNET and MORPH.

III. Approach

In this section, we describe the details of age difference dataset construction and age estimation based on the age difference.

A. Age Difference Data Collection

Training the deep age difference estimator requires face images with year labels. There are numerous resources of such images on the websites such as Filckr.com where a huge number of human photos are available with taken and

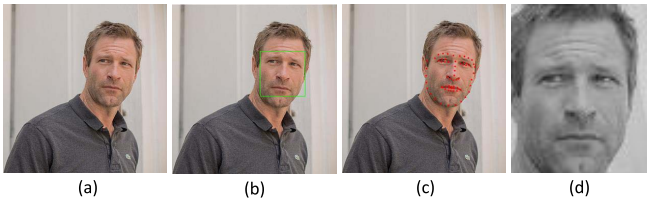


Fig. 6. The face alignment steps of face data pre-processing. (a) Original image. (b) Face detection. (c) Face landmark. (d) Face alignment.

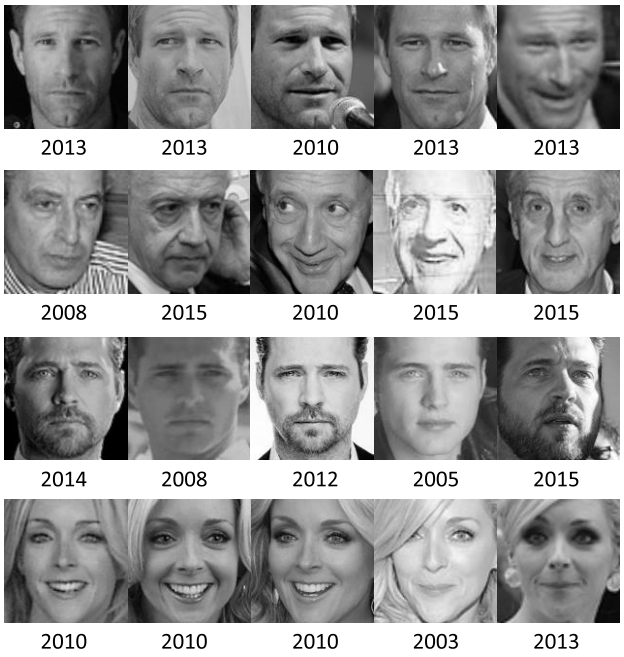


Fig. 7. Examples of normalized Flickr face dataset. Each row shows the normalized face images of the same subject attached with the photo take year.

uploaded dates. To build our dataset, we crawled millions of photos by the query names from LFW dataset. Not only the raw images, the surrounding text which contains the related information of photos such as description and taken date is also collected. Notice that during all the pre-processing steps and experiments, we always store the image data by the query names such that the images from the same subject will stay under the same path. Because only the age difference from the same subject can keep the consistency of aging influence. The age difference of different people is not in our reference.

The pre-processing of the face dataset is shown in Figure 5. It is including two flows: image processing and text processing. For the visual part, we filter all the downloaded images with face detection and alignment algorithms, as shown in Figure 6. For the text part, we extract the taken year information from surrounding text and attach it as the related image label. After the pre-processing of crawled images from websites, we contribute a human face dataset including more than 150,000 images from almost 4,000 celebrities. All the face images are normalized to 128×128. Some examples of the dataset are shown in Figure 7.

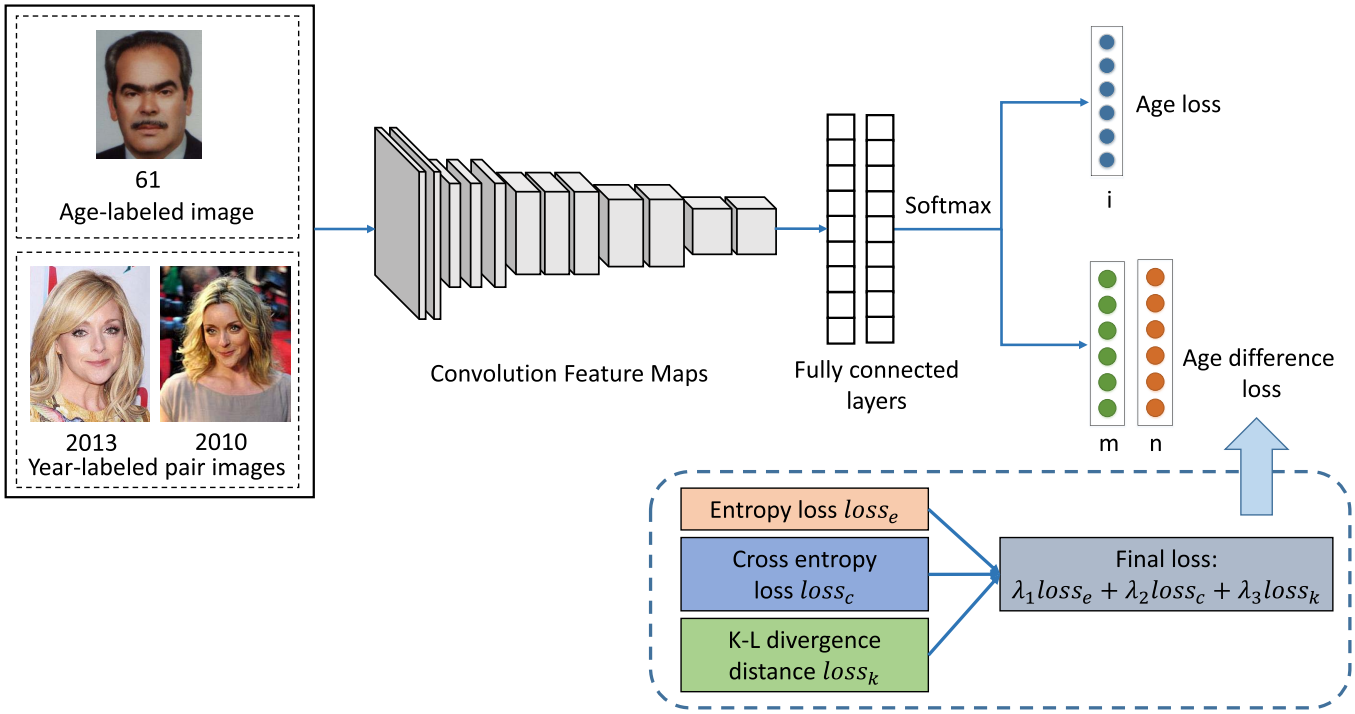


Fig. 8. An overview of the proposed deep architecture for robust age estimation.

With the face images labeled with their taken dates, we aim to explore the age information from the difference of ages. In this work, we take advantage of age difference information to improve the age estimator. Figure 8 illustrates the deep architecture of the proposed approach. We first pre-train an age estimator based on the FG-NET and MORPH aging datasets via deep CNNs with multi-label loss function. For the non-age-labeled dataset, two images from the same subject are combined as a pair. Then we fine tune the whole network with the image pairs to improve the estimator.

B. Training Objective for Age-Labeled Images

First, we train an age estimator based on the existing aging dataset. Given facial images with their ages, the age model should provide consistent estimated ages for these images. In this step, we follow the work of Geng *et al.* [2] and Geng [38] and explore the label distribution in the loss function. The advantages of label distribution, especially for age estimation task, has been demonstrated in many research works [36], [37]. All the existing aging datasets are labeled with given ages. Thus most algorithms treat the age estimation as a single label classification problem. However, human aging is generally a slow and smooth process in reality. The faces look quite similar at close ages. Geng proposed the typical label discrete distribution, e.g., Gaussian distribution, for the facial images. Label distribution not only can increase the number of labeled data but also tends to learn the similarity among the neighboring ages.

In this paper, we use Gaussian distribution to model the label distribution of ages. Let $C = \{1, 2, ..., c\}$ denote the set of possible ground truth ages. $L_m = (l_m^1, l_m^2, ..., l_m^c)$ is the label distribution for the m-th image. In this paper, we set m as the representation of an image. Given a chronological age $a \in C$, we calculate the label distribution L_m as follows. The distribution of ages $\{a-2, a-1, a, a+1, a+2\}$ is calculated as $l_m^{a_i} = l_m^a \times e^{\frac{-(a-a_i)^2}{2\theta}}$, where the Gaussian function has the mean value a_i and variance θ. For other ages, we just let $l_m^{a_i} = 0$. Finally, a normalization process is calculated to make sure that $\sum_j^c l_m^j = 1$.

In the proposed deep architecture, the Kullback-Leibler (K-L) divergences distance is set to quantify the dissimilarity between the predicted label distribution to the ground truth distribution. According to the definition of K-L divergences, the distance between two probabilities P and Q is

$$\begin{aligned} D_{KL}(P\|Q) &= \sum_i P_i \log \frac{P_i}{Q_i} \\ &= \sum_i P_i \log(Q_i) - Q_i \log(Q_i). \end{aligned} \tag{1}$$

In particular, given the training data with the Gaussian label distribution, after through the shared sub-network, an image m is mapped to a c-dimensional probability score $Q_m \in R^c$ $(Q_m^j = exp(f_m^j)/\sum_{k=1}^c exp(f_m^k))$, where f_m is the c-dimensional intermediate feature of the output of the shared sub-network for the image m and Q_m^j is the probability that image m is in age j. The loss for the image m is defined by

$$\begin{aligned} minloss &= \sum_c^{j=1} l_m^j \log(l_m^j) - l_i^j \log(Q_m^j) \\ &= \sum_c^{j=1} -l_m^j \log(Q_m^j). \end{aligned} \tag{2}$$

We optimize the network parameters via back propagation. It is a common sense that the gradient of the softmax function is

$$\frac{\partial Q_m^j}{\partial f_m^j} = Q_m^j(1 - Q_m^j). \tag{3}$$

Here we provide the gradient of $loss$ with respect to f_m^j:

$$\begin{aligned}\frac{\partial loss}{\partial f_m^j} &= \frac{\partial loss}{\partial Q_m^j} \cdot \frac{\partial Q_m^j}{\partial f_m^j} \\ &= -l_m^j \cdot \frac{1}{Q_m^j} \cdot Q_m^j(1 - Q_m^j) \\ &= Q_m^j - l_m^j. \end{aligned} \tag{4}$$

C. Training Objective for Non-Age-Labeled Images

In this step, we aim to estimate the age difference between two faces. For the images without age label, we utilize the age difference to train an age difference estimator. Given a pair of images n and m with year labels, we consider the difference of years K as the age difference. In this section, all the pair images are from the same person. Through the shared sub-network with stacked convolution layers, two images are both mapped into c-dimensional probability distributions Q_n and Q_m across C classes of ages. In order to explore the age information from the age difference, we carefully design three kinds of loss functions to leverage the age probability distributions. According to the definition of softmax, $Q_{nk} = exp(f_{nk})/\sum_{k=1}^{c} exp(f_{nk})$. f_n is the c-dimensional intermediate feature of the output of the shared sub-network for the image I_n and Q_{nk} is the probability that image n is in age k. The illustration of loss in networks is shown in Figure 8.

1) Entropy Loss: Since the output of the network is the probability distribution across a possible age range, each entry indicates the probability of the age class. Given an age probability vector, the array should have a single peak, rather than be uniformly distributed. We choose the entropy loss to satisfy this requirement. Because the entropy loss will be 0 only if one entry is 1 and all others are 0. If the probabilities are uniform values, the loss will be largest. The entropy loss for the image n is defined as

$$loss_e = -\sum_{k=1}^{c} Q_{nk} \log(Q_{nk}). \tag{5}$$

Before deriving the backward function, the gradient of Q_{nk} with respect to f_{nk} is

$$\frac{\partial Q_{nk}}{\partial f_{np}} = Q_{nk}(\delta(k = p) - Q_{np}). \tag{6}$$

The notation $\delta(k = p)$ is 1 iff $k = p$; otherwise 0. This equation is formulated according to the definition of the softmax function.

To optimize the network parameters, the gradient of $loss_e$ with respect to f_{np} is

$$\begin{aligned}\frac{\partial loss_e}{\partial f_{np}} &= \frac{\partial loss_e}{\partial Q_{nk}} \cdot \frac{\partial Q_{nk}}{\partial f_{np}} \\ &= Q_{nk}(\delta(k = p) - Q_{np}) \cdot -\sum_{k=1}^{c}(\log(Q_{nk}) + 1) \\ &= \sum_{k=1}^{c} Q_{nk}(\delta(k = p) - Q_{np}) \log(Q_{nk}) \\ &\quad + Q_{nk}(\delta(k = p) - Q_{np}) \\ &= Q_{np} \log(Q_{np}) - Q_{np} \sum_{k=1}^{c} Q_{nk} \log(Q_{nk}). \end{aligned} \tag{7}$$

2) Cross Entropy Loss: If the age difference between a pair of face images n and m is K years, assuming the image n is K years younger than the image m, then the age of image n should be no more than $c - K$ years old and the age of image m should be older than K years old. According to this, we can infer that the probability values from $c - K$ to c elements of image n should be zero and the same for image m from 0 to K elements.

Take the image n for example. We split the output of softmax layer into two parts and add up the values of elements from 0 to $c-K$ as Q_n^1 while the summation of remains is Q_n^2. This is equivalent to a binary classifier. Then we set a binary vector $b = (1, 0)$ and implement the cross entropy loss to measure the distance between the (Q_n^1, Q_n^2) and the binary vector b. The cross entropy loss for image n is defined as

$$loss_c = -\sum_{i=1}^{2} b_i \log(Q_n^i) = -\log(Q_n^1). \tag{8}$$

Here $Q_n^1 = \sum_{k=0}^{c-K} Q_{nk}$.

For the back propagation, the gradient of $loss_c$ with respect to f_{nk} is

$$\begin{aligned}\frac{\partial loss_c}{\partial f_{np}} &= \frac{\partial loss_e}{\partial Q_n^1} \cdot \frac{\partial Q_n^1}{\partial f_{np}} \\ &= -\frac{1}{\sum_{k=1}^{c-K} Q_{nk}} \left(\sum_{k=1}^{c-K} Q_{nk}(\delta(k = p) - Q_{np}) \right) \\ &= Q_{np} - \frac{Q_{nk}\delta(k = 1, \cdots, c - K)}{\sum_{k=1}^{c-K} Q_{nk}}, \end{aligned} \tag{9}$$

where $\delta(k = 1, \cdots, c - K)$ is 1 if the k is larger than 1 and smaller or equal to $c - K$ and is 0 otherwise. The output of image m is processed in the same way into (Q_m^1, Q_m^2) and compared with $b' = (0, 1)$.

3) Translation K-L Divergence Loss: Given a pair of images with age difference K of the same person, the age probability distributions should be approximate after a translation of all entries with K steps. In this step, we design a translation Kullback-Leibler (K-L) divergence loss function to quantify the dissimilarity between the distributions of image n and the translated distribution of image m.

We expect $Q_{nk} = Q'_{mk}$, $Q'_{mk} = Q_{m(k+K)}, 0 \leq k \leq c - K$ and the K-L divergences distance between these two

probabilities is defined as

$$KL(Q_n, Q'_m) = \sum_{k=1}^{c} Q_{nk} \log \frac{Q_{nk}}{Q_{m(k+K)}}. \quad (10)$$

Since K-L distance is asymmetric, we make it as symmetric as

$$loss_k = \sum_k Q_{nk} \log(\frac{Q_{nk}}{Q_{m(k+K)}}) + Q_{m(k+K)} \log(\frac{Q_{m(k+K)}}{Q_{nk}}), \quad (11)$$

and for the image m the K-L divergence loss is

$$loss_k = \sum_k Q_{n(k-K)} \log(\frac{Q_{n(k-K)}}{Q_{mk}}) + Q_{mk} \log(\frac{Q_{mk}}{Q_{n(k-K)}}). \quad (12)$$

Here the $Q_{n(k-K)}$ is the translated probability distribution of image n.

The gradient for backward for the image n is

$$\begin{aligned}\frac{\partial loss_k}{\partial f_{np}} &= \frac{\partial loss_k}{\partial Q_{nk}} \cdot \frac{\partial Q_{nk}}{\partial f_{np}} \\ &= \sum_k Q_{nk}(\delta(k=p) - Q_{np}) \log(\frac{Q_{nk}}{Q_{m(k+K)}}) \\ &\quad + Q_{nk}(\delta(k=p) - Q_{np}) \\ &\quad - \frac{Q_{m(k+K)}}{Q_{nk}} Q_{nk}(\delta(k=p) - Q_{np}) \\ &= Q_{np} \log(\frac{Q_{np}}{Q_{m(k+K)}}) - Q_{np} \sum_k Q_{nk} \log(\frac{Q_{nk}}{Q_{m(k+K)}}) \\ &\quad + Q_{np} - Q_{m(p+K)}. \end{aligned} \quad (13)$$

Finally, the overall loss of the whole age difference estimation network is

$$\min \psi = \min(\lambda_1 loss_e + \lambda_2 loss_c + \lambda_3 loss_k) \quad (14)$$

where λ_1, λ_2 and λ_3 are terms of trade-off between the errors. We set $\lambda_1 = 0.3$.

IV. Experiments

In this section, we evaluate the performance of proposed age estimation method on the public age databases and compare the results with several start-of-the-art age estimation algorithms. The following describes the details of the experiments and results.

A. Datasets Construction

For the age-labeled dataset, we evaluate the age estimation performance on two aging datasets, FG-NET and MORPH. These two datasets consist of age image sequences according to the subjects. Details of the datasets used in the experiments are summarized in Table I.

1) FG-NET [15]: There are 1,002 face images from 82 subjects in this database. Each subject has 6-18 face images at different ages. Each image is labeled with its chronological age. The ages are distributed in a wide range from 0 to 69. Besides age variation, most of the age progressive image sequences display other types of facial variations, such as significant changes in pose, illumination, expression, etc. A typical aging face sequence in this database is shown in Figure 2.

2) MORPH [16]: MORPH aging dataset is much larger than FG-NET. There are 55,132 face images from more than 13,000 subjects in this database. The average number of images per subject is 4. The ages of the face images range from 16 to 77 with a median age of 33. The faces are from different races, among which the African faces account for about 77%, the European faces account for about 19%, and the remaining 4% are Hispanic, Asian, Indian, and other races. Some typical aging faces in this database are shown in Figure 3. Although Morph has a great number of faces, most of its data seem like mug-shot images, which are quite different from the faces in the wild.

3) Year-Labeled Dataset: For the non-age-labeled dataset, we crawl the images from Flickr with the query names of LFW dataset. This dataset contains more than 150,000 face images of 3,931 subjects. The average number of images per subject is 40, which is ten times as MORPH. The range of years is from 1940 to 2014. Each image is labeled with its taken date and human ID. The max age difference is 80 years and the average of age difference is 7 years old.

TABLE I
The Illstrition of Dataset in the Experiment

Dataset	Images	Subjects	Age range
FG-Net	1,002	82	0-69
MORPH	55,134	13,618	16-77
Year-labeled	154,294	3,931	–

TABLE II
MAE (Years Old) Comparison With Different Deep CNN Architectures on the FG-NET & MORPH Dataset

Methods	MAE
AlexNet [33]	3.61
VGG-16 [40]	9.44
GoogLeNet [34]	**3.13**

B. Settings and Evaluation Measures

We implement the proposed method based on the open source `Caffe` [39] framework, which is an efficient deep neural network implementation. For the proposed method, we directly use the image pixels as input. Lots of researchers and engineers have made Caffe models for different tasks with all kinds of architectures and data. We train the age estimation model based the different CNN architectures, AlexNet [33], VGG-16 [40] and GoogLeNet [34]. The experiments results listed in Tabel II show that the GoogLeNet has the best representation of age.

The GoogLeNet is used as the shared sub-network, and the output size of the last fully connected layers is changed from 1000-dimension into c-dimension. Our networks are trained by stochastic gradient descent with 0.9 momentum and the weight decay parameter is 0.0001. All the experiments are conducted in Tesla K40c GPU with 12GB memory.

To evaluate the performance of all algorithms, we use the Mean Absolute Error (MAE) and Cumulative Score (CS) as the evaluation measures [24]. The MAE is calculated based on the average of the absolute errors between the estimated age and the ground truth (labeled age), which is represented as

$$MAE_{abs} = \frac{1}{N}\sum_{n=1}^{N}\|l_n - y_n\|, \tag{15}$$

where l_n is the ground truth label of the nth image and y_n represents the estimated age based on the proposed framework. N is the total number of testing samples. And in this work, we also consider the errors of age difference and the MAE of difference is represented as

$$MAE_{diff} = \frac{1}{N}\sum_{p,q=1}^{N}\|(l_p - l_q) - (y_p - y_q)\|, \tag{16}$$

where the l_p and l_q represent the ground truth year labels and y_p and y_q are the estimation ages for a pair of images.

The Cumulative Score (CS) is represented as

$$CS(l) = \frac{N_{e\leq l}}{N} \times 100\%, \tag{17}$$

where $N_{e\leq l}$ is the number of images with an absolute error between the estimated age and the ground truth age not greater than l years. CS is the accuracy rate for the estimation error no higher than l.

C. Testing on Age-Label Datasets

In order to verify the effectiveness of our approach, we first test the age estimation performance on age-labeled datasets. The total number of images in FG-NET and MORPH is 56,136. We randomly select 80% of images as the training data, including 44,909 images, and the remaining 11,227 images are set as the testing data. Note that there are no duplicate subjects between the training and testing sets.

We first compare our approach with traditional hand-crafted based algorithms. RED-SVM [41], k-nearest neighbors (NN), Support Vector Machine (SVM) and Support Vector Regression (SVR) are selected as the baselines. RED-SVM is a ranking-based algorithm of age estimator. SVM regards the age estimation as a single classification problem and SVR regards it as the regression problem. All these four methods are implemented by the public code LIBLinear.[1] The parameter c in RED-SVM, SVM and SVR is selected from [0.001 0.01 0.1 1 10 100 1000] and we report the best results. For kNN, we use the Euclidean distance to find the neighbors. k is set to be [1; 2; 5; 10; 20; 50; 100] and the best results are reported. For these baseline methods, we crop and align the face to the size of 64x64 pixels with the coordinates of eyes. Here we choose the BIF feature [8] because of its good performance in age estimation tasks. We also compare the results with cost-sensitive local binary feature learning (CS-LBFL) [3] because of its good performance on FG-NET and MORPH among shallow models.

[1] http://www.csie.ntu.edu.tw/cjlin/liblinear/

TABLE III
MAE COMPARISON WITH AGE ESTIMATION ALGORITHMS ON THE AGE-LABELED DATABASES

Methods	FG-NET	MORPH
RED + SVM [41]	6.33	6.57
SVR	5.45	5.52
kNN	8.13	8.22
CS-LBFL [3]	4.36	4.37
Our methods	**2.8**	**2.78**

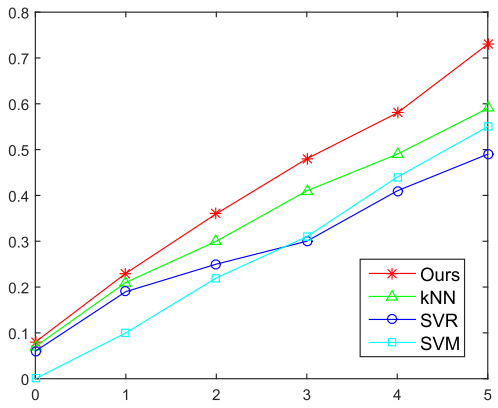


Fig. 9. The CS curve of age estimation on the datasets merged with FG-NET and MORPH.

Table III shows the comparison results of MAE on the two large datasets. As can be seen, the proposed deep-network based age estimator is significantly better than the baselines. On MORPH dataset, our method obtains MAE of 2.78, compared to RED-SVM of 6.57 and SVR of 5.52. We show the CS curves from different error levels on both datasets in Figure 9.

There are three main reasons for the good performance of our robust age estimation. First, instead of using traditional hand-crafted visual features (BIF), our method uses deep networks to learn the image representation, which increases the representational power of the image. Second, we take advantage of the age difference information. Based on the year-labeled dataset, the deep network can learn the image representation and the age difference estimation. Different from other methods which always extract the information only on standard aging datasets, our method uses a wild face dataset to guide the learning of age representation. The third is label distribution based learning. We compare the results with single age label. Different from the multi-labele learning, we choose softmax loss on the top of the network. Figure 10 shows the loss curves of single label learning and Gaussian label learning. With the same dataset and experiments setting, we can see that with the Gaussian label, the network have a faster convergence and the performance of MAE is slightly improved as listed in Table IV.

We also compare the proposed method with the most related competitors, which are also deep-networks-based methods. Table IV shows the comparison results of MAE on the MORPH2 dataset, in which the MAEs of three algorithms

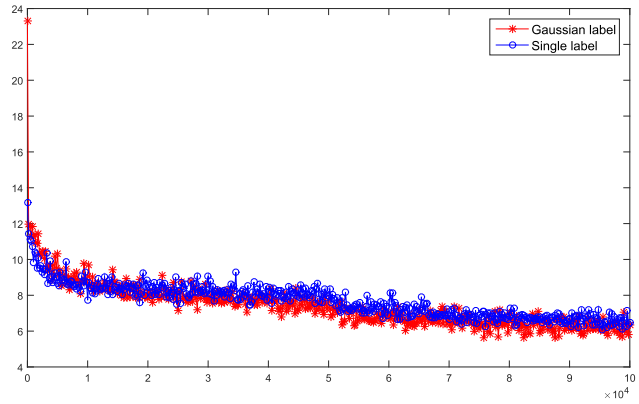


Fig. 10. The pre-trained loss curves comparison between Gaussian label and single label. The experiments is implemented on FGNET and MORPH dataset with GoogLeNet.

TABLE IV
MAE COMPARISON WITH AGE ESTIMATION ALGORITHMS ON THE YEAR-LABELED DATABASES

Methods	MORPH
GoogLeNet with year-label data	**2.78**
GoogLeNet (Gaussian label)	3.13
GoogLeNet (Single label)	3.15
Huerta [4]	3.88
Wang [14]	3.81
Wang [42]	4.77

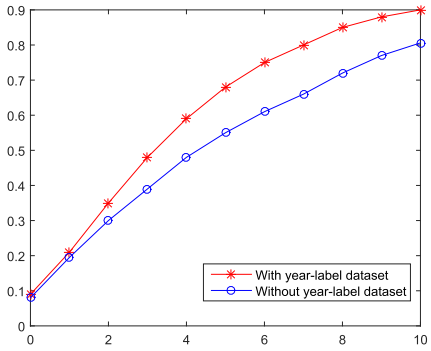


Fig. 11. The CS curve comparison between with and without year-label dataset.

are directly cited from their papers. As can be seen, the proposed age estimator with age difference assistance performs better than the baselines that also use deep learning networks. The results indicate that using the year-labeled dataset can contribute to the accuracy.

D. Testing on Year-Label Datasets

Since the year-label dataset does not have the ground truth age label, we only test the age difference estimator on it. In the same way of age-labeled data, nearly 30,000 images from 700 subjects are randomly selected as the testing set. These images are combined into 122,986 pairs. Meanwhile, more than 123,000 images from 12,300 subjects are set as training data and combined into 522,452 pairs. The MAE of the age difference is 1.74 while the average age difference of the year-labeled dataset is 7 years.

To evaluate the effectiveness of age difference information, we compare the results between with and without the year-labeled dataset. In the first two rows, we show the MAE performance. With the assistance of age difference information, the MAE on the age-labeled datasets is decreased from 3.13 to 2.78, which is the state-of-the-art result as far as we know. Figure 11 shows the comparison results of CS curve. The comparison demonstrates that the age difference information can improve the age estimator.

V. CONCLUSION

In this paper, we mainly investigate the problem of age estimation without age label and propose an approach to estimate the age of a human face with the assistance of age difference information. Given a pair of face images taken at different years of the same subjects, we exploit the age information from the images of age difference via the deep Convolutional Neural Networks (CNNs). First, we build a deep age estimator based on the standard aging datasets. A symmetric Kullback-Leibler divergence loss function is placed at the top layer of CNNs. We utilize label distribution to design the loss function. We design three kinds of loss functions on the top of the softmax layer to learn the representation of age difference. Experimental results show the advantages of the proposed age difference learning system and the state-of-the-art performance is gained.

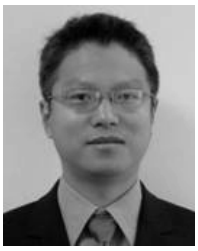
In the future work, we aim to explore more biological features of people, such as appearance, hair style, height, pose and gait.

REFERENCES

[1] A. M. Albert, K. Ricanek, and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Sci. Int.*, vol. 172, no. 1, pp. 1–9, 2007.
[2] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2401–2412, Oct. 2013.
[3] J. Lu, V. E. Liong, and J. Zhou, "Cost-sensitive local binary feature learning for facial age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5356–5368, Dec. 2015.
[4] I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, "A deep analysis on age estimation," *Pattern Recognit. Lett.*, vol. 68, pp. 239–249, Dec. 2015.
[5] K.-Y. Chang and C.-S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, Mar. 2015.
[6] H. Dibeklioglu, F. Alnajar, A. A. Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, Jun. 2015.
[7] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
[8] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 112–119.
[9] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally adjusted robust regression for human age estimation," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Apr. 2008, pp. 1–6.
[10] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[11] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3476–3483.

[12] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2014, pp. 1701–1708.
[13] M. Yang, S. Zhu, F. Lv, and K. Yu, "Correspondence driven adaptation for human profile recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2011, pp. 505–512.
[14] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jun. 2015, pp. 534–541.
[15] *The FG-Net Aging Database*, accessed on Nov. 2014. [Online]. Available: http://www.fgnet.rsunit.com/
[16] K. Ricanek, Jr., and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 341–345.
[17] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2467–2474.
[18] C. Li, Q. Liu, W. Dong, X. Zhu, J. Liu, and H. Lu, "Human age estimation based on locality and ordinal information," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2522–2534, Nov. 2015.
[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.
[20] Y. H. Kwon and N. da V. Lobo, "Age classification from facial images," *Comput. Vis. Image Understand.*, vol. 74, no. 1, pp. 1–21, 1999.
[21] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.
[22] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 34, no. 1, pp. 621–628, Jan. 2004.
[23] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[24] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, 2006, pp. 307–316.
[25] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.
[26] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, "Ranking with uncertain labels," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2007, pp. 96–99.
[27] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.
[28] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu, "Method for estimating and modeling age and gender using facial image processing," in *Proc. Int. Conf. Virtual Syst. Multimedia*, 2001, pp. 439–448.
[29] A. Günay and V. V. Nabiyev, "Automatic age classification with LBP," in *Proc. 23rd Int. Symp. Comput. Inf. Sci.*, 2008, pp. 1–4.
[30] F. Gao and H. Ai, "Face age classification on consumer images with Gabor feature and fuzzy LDA method," in *Advances in Biometrics*. Alghero, Italy: Springer, 2009, pp. 132–141.
[31] B. Ni, Z. Song, and S. Yan, "Web image mining towards universal age estimator," in *Proc. 17th ACM Int. Conf. Multimedia*, 2009, pp. 85–94.
[32] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 585–592.
[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[34] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
[35] S. Escalera *et al.*, "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 1–9.
[36] X. Liu *et al.*, "Agenet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Jun. 2015, pp. 16–24.
[37] X. Yang *et al.*, "Deep label distribution learning for apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Apr. 2015, pp. 102–108.
[38] X. Geng, "Label distribution learning," *IEEE Trans. Know. Data Eng.*, vol. 28, no. 7, pp. 1734–1748, Jul. 2016.
[39] Y. Jia *et al.* (Jun. 2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/1408.5093
[40] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[41] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "A ranking approach for human ages estimation based on face images," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 3396–3399.
[42] X. Wang and C. Kambhamettu, "Age estimation via unsupervised neural networks," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, vol. 1. Apr. 2015, pp. 1–6.

Zhenzhen Hu received the B.Sc. and Ph.D. degrees from the School of Computer and Information Science, Hefei University of Technology, in 2008 and 2014, respectively. She is currently a Research Fellow with the Cloud Computing Application and Platform Group, Nanyang Technological University, Singapore. Her research interests include computer vision and multimedia.



Yonggang Wen (S'99–M'08–SM'14) received the Ph.D. degree in electrical engineering and computer science minor in western literature from the Massachusetts Institute of Technology, Cambridge, USA, in 2008. He was with Cisco, where he led the product development in content delivery network, which had a revenue impact of three Billion U.S. dollars globally. He is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has authored over 140 papers in top journals and prestigious conferences. His research interests include cloud computing, green data center, big data analytics, multimedia network, and mobile computing. His work in multiscreen cloud social TV has been featured by global media (over 1600 news articles from over 29 countries). He received the ASEAN ICT Award 2013 (Gold Medal). His work on Cloud3DView, as the only academia entry, has received the Data Center Dynamics Awards 2015 APAC. He was a co-recipient of the 2015 IEEE Multimedia Best Paper Award. He was also a co-recipient of Best Paper Awards at the EAI/ICST Chinacom 2015, the IEEE WCSP 2014, the IEEE Globecom 2013, and the IEEE EUC 2012. He was elected as the Chair of the IEEE ComSoc Multimedia Communication Technical Committee from 2014 to 2016. He serves on editorial boards of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *IEEE Wireless Communication Magazine*, the IEEE COMMUNICATIONS SURVEY & TUTORIALS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, the IEEE ACCESS Journal, and the Elsevier *Ad Hoc Networks*.

Jianfeng Wang received the B.Eng. and Ph.D. degrees from the Department of Electronic Engineering and Information Science, University of Science and Technology of China, in 2010 and 2015, respectively. He is currently a Software Engineer with Azure Storage, Microsoft. His interests include cloud computing, machine learning, and its applications.

Meng Wang (M'09) received the B.E. and Ph.D. degrees from the Special Class for the Gifted Young and the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively. He is currently a Professor with the Hefei University of Technology, China. He has authored over 200 book chapters, journal and conference papers in these areas. His current research interests include multimedia content analysis, computer vision, and pattern recognition. He was a recipient of the ACM SIGMM Rising Star Award 2014. He is an Associate Editor of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Richang Hong (M'12) received the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2008. He was a Research Fellow with the School of Computing, National University of Singapore, from 2008 to 2010. He is currently a Professor with the Hefei University of Technology, Hefei, China. He has co-authored over 70 publications in the areas of his research interests, which include multimedia content analysis and social media. He is a member of the ACM and the Executive Committee Member of the ACM SIGMM China Chapter. He was a recipient of the Best Paper Award in the ACM Multimedia 2010, the Best Paper Award in the ACM ICMR 2015, and the Honorable Mention of the IEEE TRANSACTIONS ON MULTIMEDIA Best Paper Award. He served as the Technical Program Chair of the MMM 2016. He served as an Associate Editor of the *Information Sciences* (Elsevier) and the SIGNAL PROCESSING (Elsevier).

Shuicheng Yan (M'09—F'16) is currently a Chief Scientist with Qihoo/360 and the Deans Chair Associate Professor with the National University of Singapore. He has authored/co-authored over 100 technical papers over a wide range of research topics, with Google Scholar citation over 20,000 times and H-index 66. His research areas include machine learning, computer vision, and multimedia. He is an ISI Highly-cited Researcher of 2014, 2015, and 2016. His team received seven times winner or honorablemention prizes in PASCAL VOC and ILSVRC competitions, along with more than ten times best (student) paper prizes. He is a IAPR fellow.