2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition

Ordinal Deep Feature Learning for Facial Age Estimation

Hao Liu[1,2], Jiwen Lu[1,2,3*], Jianjiang Feng[1,2,3] and Jie Zhou[1,2,3]
[1] Department of Automation, Tsinghua University, Beijing, 100084, P.R.China
[2] Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing, 100084, P.R.China
[3] State Key Lab of Intelligent Technologies and Systems, Beijing, 100084, P.R.China
Email: h-liu14@mails.tsinghua.edu.cn; lujiwen@tsinghua.edu.cn; jfeng@tsinghua.edu.cn; jzhou@tsinghua.edu.cn.

***Abstract*— In this paper, we propose an ordinal deep feature learning (ODFL) approach for facial age estimation. Unlike conventional age estimation methods which utilize hand-crafted features, our ODFL develops deep convolutional neural networks to learn discriminative feature descriptors directly from image pixels for face representation. Motivated by the fact that age labels are chronologically correlated and age estimation is an ordinal learning computer vision problem, we enforce two criterions on the descriptors which are learned at the top of our network: 1) the topology-aware ordinal relation of face samples is preserved in the learned feature space, and 2) the age difference information of the embedded feature representation is exploited in a ranking-preserving manner. Extensive experimental results on four face aging datasets show that our approach achieves promising performance compared with the state-of-the-art methods.**

I. Introduction

Facial age estimation attempts to predict exact age values for given facial images, which plays an important role in the human-computer interaction, visual advertisements and biometrics [1]–[3]. While extensive efforts have been devoted, facial age estimation still remains a challenging problem due to two aspects: 1) large variations caused by cluttered occlusions, facial poses and expressions, and 2) aging labels have chronological ordinal relation.

Existing facial age estimation systems usually consists of two key components: extracting face features [2], [4]–[6] and learning age estimators [7], [8]. However, most features employed in these methods are hand-crafted, which requires strong prior knowledge by hand. To address this, learning-based feature representation methods [9]–[11] have been proposed to learn discriminative feature representation directly from raw pixels. For example, Fu *et al.* [10] proposed a holistic feature learning method by leveraging a discriminative manifold learning technique. Lu *et al.* [11] addressed the cost-sensitive problem for age estimation by learning local binary codes for face representation. However, their methods utilize linear feature filters so that they are not powerful enough to exploit the complex and nonlinear relationship between face samples and age labels. To address this nonlinear issue, deep learning methods [12]–[16] have been applied to model the relationship between face features and aging process by a series of nonlinear transformations. For example, Yi *et al.* [12] employed multi-scale deep feature representation via convolutional neural networks (CNN) to predict the age value with the additional gender and ethnicity information. Niu *et al.* [16] developed an ordinal ranker with multiple binary classifiers under the CNN architecture.

Unlike existing deep learning-based facial age estimation methods which ignore to explicitly consider the structural order relationship among face images (*i.e.*, quadruplet and triplet based comparisons), we propose an ordinal deep feature learning method, dubbed ODFL, to learn age-adaptive face descriptors with CNN to exploit the topology-aware ordinal relation for face presentation. To achieve this, we enforce two important criterions at the top of our network and optimize the parameters of the network with back-propagation. We conduct experiments on four face aging datasets and obtain significant performance in comparisons with the state-of-the-art facial age estimation methods.

The rest of this paper is organized as follows: Section II briefly reviews some related work. Section III describes the proposed ordinal deep feature learning method for facial age estimation in details. Section IV reports experimental results and analysis, and Section V concludes the paper.

II. Related Work

A. Facial Age Estimation

Numerous facial age estimation methods [8], [17]–[21] have been proposed over the past two decades. For example, Lanitis *et al.* [17] applied an age regression method to address the face aging problem. Zhang and Yueng [18] proposed an age estimation method by using a multi-task Gaussian process (MTWGP). Chang *et al.* [8] presented an ordinal hyperplane ranking (OHRanker) method which divided the age estimation problem as a series of sub-problems of binary classifications. Geng *et al.* [20] proposed a label distribution learning (LDL) approach to model the relationship between face images and age labels. However, these methods usually employ hand-crafted features such as the holistic subspace feature [9], [22], local binary pattern (LBP) [5] and the bio-inspired feature (BIF) [2] for face representation, which requires strong expert knowledge by hand. To address this, several attempts have been made to learn discriminative face descriptors by using advanced feature learning approaches [11], [21], [23], [24]. For example, Guo *et al.* [24] proposed a holistic feature learning approach by utilizing a manifold learning technique. Lu *et*

* Corresponding author.

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.28

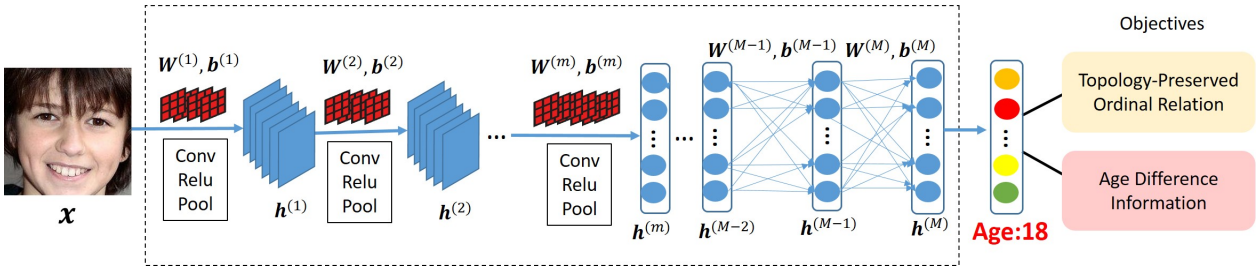

Fig. 1. The framework of the proposed ODFL. During the training stage, we enforce two objectives on learning age-related face descriptors to exploit both the topology-preserving ordinal relation and age difference information at the top layer of the designed network, and the parameters of the network are optimized via back-propagation. During the testing stage, we directly feed the face image to the trained network. Having obtained the face representation, we put it to a learned age estimator and obtain the exact age value.

al. [11] proposed a local binary feature learning method (CS-LBFL) to learn a face descriptor which is robust to local illumination. Nevertheless, these methods aim to seek simple feature filters, so that they are not powerful enough to exploit the nonlinear relationship of face samples in such cases that facial images are exposed to large variations of diverse facial expressions and ordinal aging labels.

B. Deep Learning

Recently, deep learning methods have been applied to many facial analysis tasks including face detection [25], face alignment [26] and face recognition [27], [28]. For example, Zhang *et al.* [26] utilized stacked auto-encoder networks to estimate facial landmarks in a coarse-to-fine manner. Sun *et al.* [27] developed a DeepID2 network to reduce the personalized inter-covariance jointly by using the identification and verification signals. Parkhi *et al.* [28] proposed a VGG Face Net with a very deep architecture, which was pretrained by a large scale face dataset for face recognition. Inspired by the aforementioned works which learn task-adaptive face feature representation, deep learning has been also used to learn a set of nonlinear feature transformations for facial age estimation [13], [16], [29]–[32]. For example, Levi *et al.* [32] proposed a multi-task method with CNN to jointly address the age and gender classification in a unified framework. Yang *et al.* [33] employed deep scattering transform networks (DeepRank) to predict ages via category-wise rankers. Niu *et al.* [16] developed a CNN-based ordinal regression (OR-CNN) method with multiple binary outputs for age prediction. While significant performance can be obtained under these methods, they ignored to take advantages of the quadruplet-based ordinal relation during batch-wise training procedure in deep learning, which makes the learned features less efficiency for age prediction.

In contrast to previous approaches, our proposed ODFL learns face descriptors directly from image pixels with CNN by exploiting the structural and high-order information including both quadruplet and triplet ordinal relations. We show that our method achieves better performance than the state-of-the-art facial age estimation methods on four face aging datasets.

III. Proposed Method

Conventional facial age estimation methods [8], [17], [18], [20] utilize hand-crafted features, which may loss some crucial information. Learning-based face representation methods [10], [11], [23] learn linear feature filters which are not powerful enough to model the nonlinear relationship of face data and age labels. To address both the nonlinear and feature learning issues, deep learning [13], [29]–[31], [34] has been adopted to learn discriminative features from raw pixels under the CNN architecture. However, these methods cannot directly model the topo-structure and high-order relations across age labels for real-world aging pattern. To address this, we introduce an ordinal deep feature learning (ODFL) approach to learn face descriptors for facial age estimation. Specifically, our ODFL aims to learn a series of hierarchical nonlinear transformations by enforcing two importance objectives, where both the topology-preserving ordinal relation and age difference information are exploited in the learned face descriptors. In this section, we will describe the proposed learning criterions and detail the optimization procedure.

A. Ordinal Deep Feature Learning

Fig. 1 shows the framework of the proposed method. Let $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote the training set which contains N samples, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the ith face of d pixels. The goal of our model aims to learn to compute feature representation $f(\mathbf{x}_i)$ with CNN for the ith face image $\mathbf{x}_i$. We feed the face image to the designed CNN and obtain the immediate feature representation, which is computed as follows:

$$f(\mathbf{x}_i) = \mathbf{h}_i^{(m)} = \text{pool}\left(\text{ReLU}(\mathbf{W}^{(m)} \otimes \mathbf{x}_i + \mathbf{b}^{(m)})\right), \quad (1)$$

where pool$(\cdot)$ denotes the max pooling operation, ReLU$(\cdot)$ denotes the nonlinear *ReLU* function, and m is specified to a set of $\{1, 2, \cdots, M-2\}$ which represents the mth layer.

To learn efficient face descriptors for facial age estimation, the key design lies in preserving the ordinal relation among training samples in the transformed feature space. To achieve this, we define the training loss by including two terms: the topology-preserving ordinal relation term J_1 and the age difference information term J_2 at the top of our network.

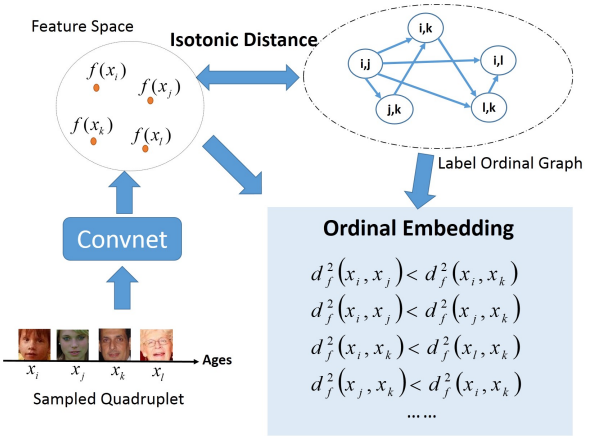

Fig. 2. Topology-Preserving Ordinal Relation. Given a quadruplet of face samples and age labels from a training batch, we construct a directed unweighted topology as the label ordinal graph towards ordinal embedding. Our ODFL aims to learn a deep convnet, where the topology-aware ordinal relation within the label ordinal graph has isotonic distance with that in the learned feature space. As a result, the topology-preserving ordinal relation is preserved in the obtained face descriptors which are computed by the trained convnet.

The parameters of the network are optimized via back-propagation.

The face descriptor at the top layer of our network is computed as follows:

$$f(\mathbf{x}_i) = \mathbf{h}_i^{(M)} = \sigma(\mathbf{W}^{(M)}\mathbf{x}_i + \mathbf{b}^{(M)}), \tag{2}$$

where $\mathbf{W}^{(M)}$ and $\mathbf{b}^{(M)}$ denote the weights and bias of the top layer, respectively, and $\sigma(\cdot)$ denotes the nonlinear function.

To sum up the total weights, we collect $\{1, 2, \cdots, M\}$ for m to train the whole CNN based on the dissimilarity on the face pair of $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$, which is computed as follows:

$$d_f^2(\mathbf{x}_i, \mathbf{x}_j) = \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2, \tag{3}$$

where $\|\cdot\|_2$ denotes the Euclidean distance in the learned feature space.

Topology-Preserving Ordinal Relation. Unlike conventional facial age estimation methods [11], [16], [19] which learn age rankers based on pairwise comparisons, we construct a label ordinal graph based on a set of quadruplets from training batches, and the defined objective enforces that the ordinal relation in the learned feature space should be isotonic to that in the label space [35]. The goal of our ODFL is to map the face samples to a latent space, where the topology-aware ordinal relation is preserved in the learned face descriptors according to the distance of age labels. To better measure the distance between pairs of age labels, we introduce a label embedding method [36] to smooth the label distance.

As illustrated in Fig. 2, suppose we have a sampled quadruplet (i, j, k, l) from the training batch $\mathscr{B}$ with the knowing age labels (y_i, y_j, y_k, y_l). Based on the age labels, our model encodes such a quadruplet with a particular subset of ordinal constraints as follows:

$$\delta(y_i, y_j) < \delta(y_k, y_l), \forall (i, j, k, l) \subseteq \mathscr{B}, \tag{4}$$

where $\delta(\cdot, \cdot)$ denotes the smooth function, which is viewed as a dissimilarity degree between a pair of age labels and is defined by the Gaussian function as follows:

$$\delta(y_i, y_j) = \delta_{ij} = \exp^{\frac{-(y_i - y_j)^2}{H^2}}, \tag{5}$$

where H denotes the label difference threshold to determine the variance of age label distribution.

To model the topo-structure for the quadruplet of age labels, we construct a label graph $G = (V, E) = [n]^4$, where each node $\delta_{ij} \in V$ represents the age dissimilarity degree between the ith and jth samples, while each directed edge $e_{(i,j,k,l)\subseteq\mathscr{B}} \subseteq E$ represents an ordinal relation of $\delta_{ij} < \delta_{kl}$. Our ODFL aims to encode items in $\mathscr{B}$ as projected feature representation such that the ordinal constraints are preserved by an isotonic distance, which is defined as follows:

$$\delta_{ij} < \delta_{kl} \Longrightarrow d_f^2(\mathbf{x}_i, \mathbf{x}_j) < d_f^2(\mathbf{x}_k, \mathbf{x}_l), \tag{6}$$

which means the topology-aware ordinal relation within the label ordinal graph has the isotonic distance with that in the learned feature space (see more details in Fig. 2). There are two common situations for (6), i.e., quadruplet ordinal relation where $(i, j, k, l) \subseteq \mathscr{B} \subseteq [n]^4$ and $(i, j, i, k) \subseteq \mathscr{B} \subseteq [n]^3$. Hence, the objective takes advantages of the fully structural ordinal relation of training batches, so that the high-order quadruplet and triplet based comparisons can be taken into account in the feature space simultaneously, where the distance of the face pair of the ith and jth samples should be smaller than that with the face pair of the kth and lth samples.

To involve the label information, we utilize the constructed ordinal label graph G to train the designed network in a globally supervised manner. For the ordinal relation of $e_{(i,j,k,l)\subseteq\mathscr{B}} \subseteq E$ in the batch $\mathscr{B}$, we expect the relation of age dissimilarity degree should be preserved by the learned feature space under the constraint of (6). To achieve this, we employ the hinge loss to optimize the violates of unsatisfied quadruplet comparisons. Hence, the objective J_1 is formulated as follows:

$$\sum_{v_{ij}, v_{kl} \in G} \zeta(v_{ij}, v_{kl}) \cdot \max[0, \alpha - d_f^2(\mathbf{x}_i, \mathbf{x}_j) + d_f^2(\mathbf{x}_k, \mathbf{x}_l)], \tag{7}$$

where $\zeta(v_{ij}, v_{kl})$ indicates 1 if there is a vertex v_{ij} to v_{kl}, and 0 otherwise, and α denotes a thresholding margin which is set to 1.

Age Difference Information. To better improve the discriminativeness of the face descriptors, we introduce a weighted ranking approximation method [37] to consider the ranking-preserving age difference information for the sampled triplets based on age gaps. As demonstrated in Fig. 3, we define an objective to measure the age difference information. Specifically, given a triplet of an anchor sample and other two samples, based on this anchor sample, the objective enforces that the difference of a pair with a small

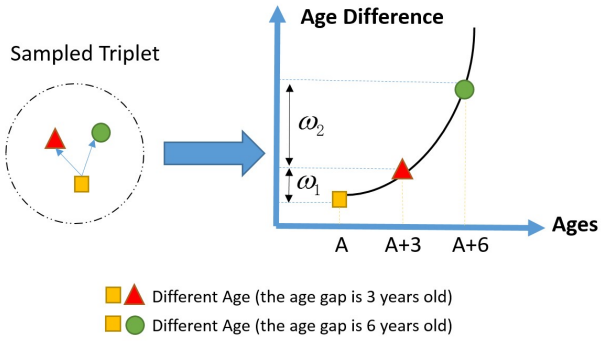

Fig. 3. Age Difference Information. Suppose there are three face samples from the training set and let the yellow square denote the anchor sample. Based on the anchor sample, the red triangle represents the face sample with an age gap of 3 years old and the green circle denotes that with an larger age gap of 6 years old. Our ODFL aims to learn a set of nonlinear feature transformations, where a face pair with a larger age gap has a larger ranking weight ω_2 than the ranking weight ω_1 with a smaller age gap. As a result, the ranking-preserving age difference information can be exploited in the learned feature space to reinforce our model (best viewed in color pdf file).

age gap should be smaller than that of a pair with a large age gap in the learned feature space. To this end, the age difference is weighted dynamically in the embedded feature space according to different age gaps, and the ranking weights are computed to show how they exploit different relations for different age gaps. Therefore, our goal of J_2 is to minimize the following objective function:

$$\sum_{p}^{P}\left(1-\ell_{p1,p2}(\tau-d_f^2(\mathbf{x}_{p1},\mathbf{x}_{p2}))\cdot\omega_{y_{p1},y_{p2}}\right) \tag{8}$$

where $(p1,p2)$ denotes a face pair with different age gaps for a given anchored face sample p. $\ell(p1,p2)$ denotes the indicator which is set to 1 if the face pair belongs to the same age labels, and is set to -1 otherwise. τ represents a threshold between the distance of the face pair with larger age gaps and that of the face pair with smaller age gaps (τ is specified to 1 in this work). y_{p1} and y_{p2} represent the age gaps computed based on the ground-truth. $\omega_{y_{p1},y_{p2}}$ denotes the smoothness weighting function, which is computed as follows:

$$\omega_{y_{p1},y_{p2}}=\begin{cases}(|y_{p1}-y_{p2}|+1)^{\eta}, & \text{if} \quad y_{p1}\neq y_{p2}\\ 1, & \text{otherwise}\end{cases} \tag{9}$$

where η is a constant parameter that describes the tolerance level of varying age relationship.

With the defined age-difference specific objective, the ranking weights are preserved by the smooth function instead of treating all pairs with different age gaps equally, where the chronological aging process can be well measured in the embedded feature space. In this way, the face representation is embedded to exploit the age difference information based on the sampled face pairs to boost the facial age estimation performance.

B. Formulation

To combine both topology-preserving ordinal relation and age difference information in our training loss, we formulate the following objective function:

$$\begin{aligned}\min_{\mathbf{W}} J &= J_1+\lambda_1 J_2+\lambda_2 J_3 = \\ &\sum_{v_{ij},v_{kl}\in G}\zeta(v_{ij},v_{kl})\cdot\max[0,\alpha-d_f^2(\mathbf{x}_i,\mathbf{x}_j)+d_f^2(\mathbf{x}_k,\mathbf{x}_l)] \\ &+\lambda_1\sum_{p}^{P}\left(1-\ell_{p1,p2}(\tau-d_f^2(\mathbf{x}_{p1},\mathbf{x}_{p2}))\cdot\omega_{y_{p1},y_{p2}}\right) \\ &+\lambda_2\sum_{m=1}^{M}(\|\mathbf{W}^{(m)}\|_F^2+\|\mathbf{b}^{(m)}\|_2^2),\end{aligned} \tag{10}$$

where λ_1 and λ_2 are the hyper-parameters to balance these terms and $\|\mathbf{W}^{(m)}\|_F^2$ denotes the Frobenius norm of matrix $\mathbf{W}^{(m)}$ to prevent the parameters of deep network from overfitting, respectively.

The first term J_1 in (10) is to preserve the topology-preserving ordinal relation for each sampled quadruplet. Moreover, the fully order relationship of both quadruplet and triplet ranking comparisons can be preserved in the learned feature space in a purely supervised way. The second term J_2 in (10) attempts to dynamically assign the ranking-preserving weights to achieve the age difference information for triplets according to age gaps, where the ranking-preserving order relationship is exploited across age labels. The third term J_3 enforces the regularization on network parameters to reduce the model complexity.

C. Optimization

To optimize J_1 in (10), we present a landmark-based ordinal embedding method (LOE) [38], which considers the triplet comparisons from any training samples to the landmark. In this way, the number of ordinal constraints reduces from n^4 to $n\cdot L^2$, where L denotes the landmark number. Moreover, we apply a logistic loss function to relax the maximum non-convex function $\max[0,\Psi]$ that is not easy to optimize by $g(\Psi)=\frac{1}{\beta}\log(1+\exp(\beta\Psi))$, where β is a sharpness parameter. Hence, the formulation of (10) is written as follows:

$$\begin{aligned}\min_{\mathbf{W}} J &= J_1+\lambda_1 J_2+\lambda_2 J_3 \\ &=\sum_{i=1}^{n}\sum_{j,k=1}^{L}\zeta(v_{ij},v_{ik})\cdot g(\alpha-d_f^2(\mathbf{x}_i,\mathbf{x}_j)+d_f^2(\mathbf{x}_k,\mathbf{x}_l)) \\ &+\lambda_1\sum_{p}^{P}(1-\ell_{p1,p2}(\tau-d_f^2(\mathbf{x}_{p1},\mathbf{x}_{p2}))\cdot\omega_{y_{p1},y_{p2}}) \\ &+\lambda_2\sum_{m=1}^{M}(\|\mathbf{W}^{(m)}\|_F^2+\|\mathbf{b}^{(m)}\|_2^2),\end{aligned} \tag{11}$$

where τ denotes the threshold and is set to 1.

To solve the relaxed optimization problem of (11), we leverage the stochastic gradient descent method to obtain the

Algorithm 1: The Optimization Procedure of ODFL.

Input: Training set: $X = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, learning rate ρ and iteration number T.

Output: The network parameters $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{i=1}^{M}$.

Step 1 (Parameters Intialization): Initialize the parameters $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{i=1}^{M}$ by the pretrained network.

Step 2 (Optimization via Back-Propagation):

repeat

* Randomly select an quadruplet (i, j, k, l) from a training batch $\mathscr{B}$, and then construct the label ordinal graph G by using the label quadruplet (y_i, y_j, y_k, y_l) according to (4) .
* Perform forward propagation and map G to a landmark-based graph based on LOE [38].
* Perform backward propagation and compute the gradients
* Update the parameters according to (12) and (13).

until *convergence or reaching the maximum iteration number* T;

Return $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{i=1}^{M}$.

gradients of the parameters $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}$ w.r.t. the objective function J, where $m = \{1, 2, ..., M\}$.

Then, $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$ are updated by using the gradient algorithm as follows until convergence:

$$\mathbf{W}^{(m)} = \mathbf{W}^{(m)} - \rho \frac{\partial J}{\partial \mathbf{W}^{(m)}}, \tag{12}$$

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m)} - \rho \frac{\partial J}{\partial \mathbf{b}^{(m)}}, \tag{13}$$

where ρ denotes the learning rate, which controls the convergence speed of the objective function J. **Algorithm 1** details the training procedure of the proposed ODFL.

IV. Experiments

We evaluated our ODFL on four widely benchmarking datasets including the MORPH (Album2) [39], FG-NET [17], FACES [40] and the apparent facial age estimation [41] datasets. The followings describe the details of our experimental settings and results.

A. Experimental Settings and Implementational Details

We detected the face bounding boxes on the original face images based on the open source computer vision library DLIB [42]. For each facial image, we detected three landmarks including two centers of eyes and the nose base to align the face into the canonical coordinate system by using the similar transformation. The aligned faces were fed to the designed network and then the exact age values were predicted by the learned OHRanker [8]. For the parameters employed in our ODFL, we set $H = 5$, $\eta = 0.5$, $\lambda_1 = 0.3$ and $\lambda_2 = 0.001$ by cross-validation. For the parameters of the designed network, we specified the values of the weight decay, moment and learning rate empirically to 0.0001, 0.9 and 0.001, respectively. The whole training procedure of the network converged in 5 iterations.

B. Evaluation Metrics

For the evaluation metrics, we utilized the mean absolute error (MAE) [1], [9] to measure the error between the predicted age and the ground-truth, which is computed as follows:

$$\epsilon = \frac{\|\hat{y} - y^*\|_2}{N} \tag{14}$$

where $\hat{y}$ and y^* denote predicted and ground-truth age value, respectively, and N denotes the number of the testing facial images.

We also applied the cumulative score (CS) [18], [21], [22], [33] curves to quantitatively evaluate the performance of age estimation methods. The cumulative prediction accuracy at the error ϵ is computed as:

$$CS(\theta) = \frac{N_{\epsilon \leq \theta}}{N} \times 100\% \tag{15}$$

where $N_{\epsilon \leq \theta}$ is the number of images on which the error θ is no less than ϵ.

C. Experiments on the MORPH dataset

The MORPH (Album 2) dataset [39] consists of 55608 face images from about 13000 subjects. The age range lies from 16 to 77 years old and there exists averaging 4 samples per person. We performed 10-folds cross-validation for evaluation by following the settings in [11].

1) Comparisons with the State-of-the-art Methods: We compared our model with several different state-of-the-art facial age estimation approaches. We created a baseline method by utilizing the bio-inspired feature (BIF) [2] and KNN, and implemented the state-of-the-art methods including OHRanker [8] and CS-LBFL [11] by following the details from the original papers. Table I tabulates the MAEs, where the MAEs of the state-of-the-arts were directly cropped from the related papers. Fig. 4 shows the CS curves. According to the results, we see that our ODFL outperforms the state-of-the-art methods and even obtains better performance than that of the deep learning methods such as DeepRank [33] and OR-CNN [16].

2) Comparisons with Different Deep Learning Methods: We also compared our ODFL with different deep learning methods. To be specific, we first employed the pretrained VGG Face Net [28] without the fine-tuning training as the feature extractors. We created a baseline method with the unsupervised VGG features and KNN. Then, we deployed the softmax loss [43] as the single label method, and the deep label distribution learning [14] as the Gaussian label methods at the top of the VGG Face Net and finetuned the network. Table II tabulates the performance of different deep learning methods. We see that our model obtains the best performance, which is because the structural ordinal relation is exploited by our model in the learned face feature representation, which take advantages of the fully order relationship of quadruplet comparisons.

TABLE I

COMPARISON OF MAES WITH DIFFERENT STATE-OF-THE-ART APPROACHES ON THE MORPH DATASET.

Method	MAE
BIF+KNN	9.64
AGES [1]	8.83
Raw+OHRanker	7.34
LBP+OHRanker	6.88
BIF+OHRanker	6.49
MTWGP [18]	6.28
LDL [20]	5.69
CPNN [20]	5.67
CA-SVR [44]	4.87
BIF+OLPP [45]	4.20
CS-LBFL [11]	4.52
CS-LBMFL [11]	4.37
DeepRank [33]	3.57
DeepRank+ [33]	3.49
OR-CNN [16]	3.27
ODFL	**3.12**

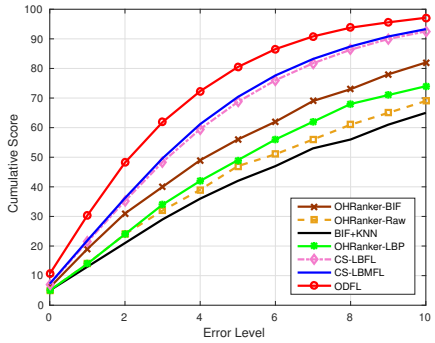

Fig. 4. The CS curves of our ODFL compared with different facial age estimation methods on the MORPH dataset.

3) Comparisons with Existing Networks: We compared the performance of our ODFL with existing deep networks including AlexNet [43], ResNet [46] and VGG Face Net [28]. Specifically, we directly deployed our proposed objectives at the top of the pretrained networks and finetuned them. Note that the ResNet and VGG Face Net were fed with the faces in the size of 224×224 and 227×227 for the AlexNet. Table III tabulates the results of our ODFL compared with different deep networks. According to the results, we see that our ODFL with the VGG Face Net obtains the best performance. It is because the VGG Face Net were pretrained by a large amount of face images for 2622 person identities, which achieves to capture more facial patterns than those of any other networks and further improves the capacity of the learned face features for age.

4) Computational Time: Our ODFL was implemented by the open source Caffe [47] deep learning toolbox. We trained our model with a speed-up parallel computing technique by using single GPU with NVIDIA GTX 970. Table III tabulates the comparisons of the computational time during the testing phase. We see that the VGG Face Net proceeds averaging

TABLE II

COMPARISON OF MAES WITH DIFFERENT DEEP LEARNING APPROACHES ON THE MORPH DATASET.

Method	MAE
unsupervised VGG + KNN	7.21
unsupervised VGG + OHRanker	4.58
VGG + Single Label	3.63
VGG + Gaussian Label	3.44
ODFL	**3.12**

TABLE III

COMPARISON OF MAES AND COMPUTATION TIME OF OUR ODFL COMPARED WITH DIFFERENT DEEP NETWORK ARCHITECTURES ON THE MORPH DATASET.

Method	MAE	Testing Time
AlexNet [43]	3.72	2425.3 imgs/s
ResNet [46]	3.47	256.8 imgs/s
VGG Face Net [28]	**3.12**	143.2 imgs/s

143.2 images per second with single GPU. Moreover, the OHRanker employed in our experiments takes 0.04 seconds by using an Intel i7-CPU@3.40GHz PC, which satisfies the real-time requirement.

D. Experiments on the FG-NET dataset

There are 1002 images from 82 persons in FG-NET dataset [17] and there exists averaging 12 samples for each person. The age range covers from 0 to 69. To conduct the age estimation experiments, we employed the leave-one-person-out (LOPO) evaluation. Specifically, we randomly selected face images from one person as testing images, and the remaining were used for training. Table IV and Fig. 5 shows the MAEs and the CS curves of our ODFL compared with the state-of-the-arts, respectively. From the results, we see that our ODFL outperforms of the state-of-the-arts.

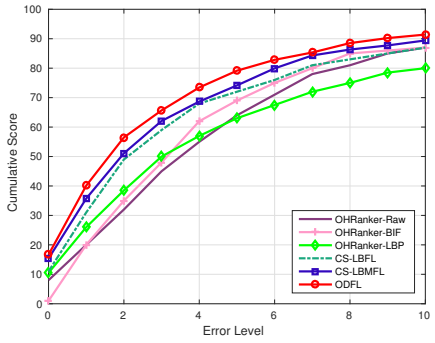

Fig. 5. The CS curves of our ODFL compared with different facial age estimation methods on the FG-NET dataset.

TABLE IV

COMPARISON OF MAES COMPARED WITH STATE-OF-THE-ART APPROACHES ON THE FG-NET DATASET.

Method	MAE
BIF+KNN	8.24
Raw+OHRanker	6.25
LBP+OHRanker	4.92
BIF+OHRanker	4.48
RUN [48]	5.78
AGES [1]	6.77
MTWGP [18]	4.83
PLO [6]	4.82
LDL [20]	5.77
CPNN [20]	4.76
CA-SVR [44]	4.67
CS-LBFL [11]	4.43
CS-LBMFL [11]	4.36
ODFL	**3.89**

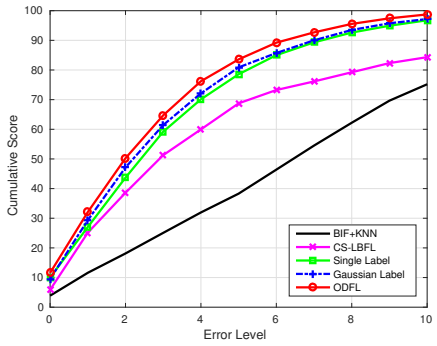

Fig. 6. The CS curves of our ODFL compared with different facial age estimation methods on the apparent facial age estimation dataset.

E. Experiments on the Apparent Age Estimation dataset

We have also investigated our method on the apparent age estimation dataset [41]. This dataset contains 4112 images for training and 1500 images for validation. The age range covers from 0 to 100 years old. To conduct the experiments of our ODFL, we also created the single label and Gaussian label methods with the VGG Face Net. Table V tabulates the MAEs and Gaussian errors [41] and Fig. 6 shows the CS curves, respectively. From these results, we see that our ODFL performs better than other deep learning methods without any additional labeled data. Furthermore, we illustrated some resulting samples in Fig. 7, where the

Fig. 7. The selected examples from the apparent age estimation dataset, where the age prediction errors are below one year old. According to these resulting samples, we see that our ODFL is robust to large variances of facial wearing glasses, poses and expressions.

TABLE V

COMPARISON OF MAES AND GAUSSIAN ERRORS WITH DIFFERENT FACIAL AGE ESTIMATION APPROACHES ON THE APPARENT AGE ESTIMATION DATASET.

Method	MAE	Gaussian Error
BIF+KNN	7.19	0.620
CS-LBFL	5.12	0.422
Single Label	4.58	0.416
Gaussian Label	4.31	0.363
ODFL	**4.12**	**0.339**

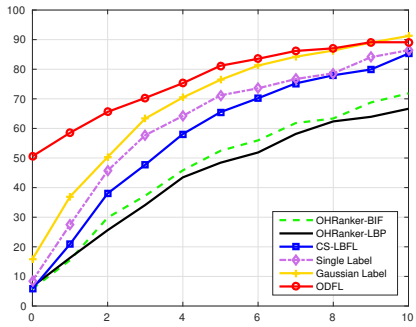

Fig. 8. The CS curves of our ODFL compared with different facial age estimation methods for Happy Expression on the FACES dataset.

age prediction errors are below one year old. From these samples, we see that our ODFL are robust to large variations caused by diverse facial expressions and aspect ratios.

F. Experiments on the FACES dataset

The FACES dataset [40] contains 2052 face images from 171 persons. The age range covers from 19 to 80 years old. For each person, there are six expressions including neutral, sad, disgust, fear, angry and happy. In our experimental setting, we conducted the experiments under the same expression. Table VI tabulates the MAEs and Fig. 8 shows the CS curves compared with different facial age estimation approaches, respectively. According to the results, we see the our ODFL obtains significant performance compared with any other state-of-the-art methods. This is because our method achieves the age-adaptive information across different facial expressions based on the VGG Face Net, which contributes to the improvements for facial age estimation on this dataset.

V. CONCLUSIONS AND FUTURE WORK

We have proposed a new feature learning method called ordinal deep feature learning for facial age estimation. Experimental results on four datasets show the effectiveness of the proposed method. It is desirable to address facial age estimation with the feed-back networks to further exploit with the complementary information for the personalized aging pattern in the future work.

TABLE VI

COMPARISON OF MAEs WITH DIFFERENT STATE-OF-THE-ART APPROACHES ON THE FACES DATASET.

Method	Neutral	Happy	Disgust	Fearful	Sad	Angry
LBP+OHRanker	5.16	7.64	8.31	7.00	6.87	7.87
BIF+OHRanker	6.36	8.88	9.20	7.30	9.09	8.86
CS-LBFL [11]	5.06	6.53	7.15	6.32	6.27	6.94
DeepRank [33]	5.99	7.12	8.15	6.35	7.77	6.68
ODFL	**3.48**	**3.52**	**4.41**	**4.52**	**3.96**	**3.87**

ACKNOWLEDGEMENT

This work is supported by the National Key Research and Development Program of China under Grant 2016YFB1001001, the National Natural Science Foundation of China under Grants 61672306, 61225008, 61572271, 61527808, 61373074 and 61373090, the National 1000 Young Talents Plan Program, the National Basic Research Program of China under Grant 2014CB349304, the Ministry of Education of China under Grant 20120002110033, and the Tsinghua University Initiative Scientific Research Program.

REFERENCES

[1] Geng, X., Zhou, Z., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. PAMI **29**(12) (2007) 2234–2240

[2] Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: CVPR. (2009) 112–119

[3] Shu, X., Tang, J., Lai, H., Liu, L., Yan, S.: Personalized age progression with aging dictionary. In: ICCV. (2015) 3970–3978

[4] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. TPAMI **23**(6) (2001) 681–685

[5] Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. PAMI **28**(12) (2006) 2037–2041

[6] Li, C., Liu, Q., Liu, J., Lu, H.: Learning ordinal discriminative features for age estimation. In: CVPR. (2012) 2570–2577

[7] Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. TMM **10**(4) (2008) 578–584

[8] Chang, K., Chen, C., Hung, Y.: Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: CVPR. (2011) 585–592

[9] Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. TMM **10**(4) (2008) 578–584

[10] Fu, Y., Guo, G., Huang, T.S.: Age synthesis and estimation via faces: A survey. PAMI **32**(11) (2010) 1955–1976

[11] Lu, J., Liong, V.E., Zhou, J.: Cost-sensitive local binary feature learning for facial age estimation. TIP **24**(12) (2015) 5356–5368

[12] Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. In: ACCV. (2014) 144–158

[13] Liu, X., Li, S., Kan, M., Zhang, J., Wu, S., Liu, W., Han, H., Shan, S., Chen, X.: Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In: ICCVW. (2015) 258–266

[14] Yang, X., Gao, B., Xing, C., Huo, Z., Wei, X., Zhou, Y., Wu, J., Geng, X.: Deep label distribution learning for apparent age estimation. In: ICCVW. (2015) 344–350

[15] Wang, X., Guo, R., Kambhamettu, C.: Deeply-learned feature for age estimation. In: WACV. (2015) 534–541

[16] Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output cnn for age estimation. In: CVPR. (2016) 4920–4928

[17] Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. PAMI **24**(4) (2002) 442–455

[18] Zhang, Y., Yeung, D.: Multi-task warped gaussian process for personalized age estimation. In: CVPR. (2010) 2622–2629

[19] Chang, K., Chen, C., Hung, Y.: A ranking approach for human ages estimation based on face images. In: ICPR. (2010) 3396–3399

[20] Geng, X., Yin, C., Zhou, Z.: Facial age estimation by learning from label distributions. PAMI **35**(10) (2013) 2401–2412

[21] Guo, G., Mu, G.: Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: CVPR. (2011) 657–664

[22] Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. TIP **17**(7) (2008) 1178–1188

[23] Liu, H., Ji, R., Wu, Y., Liu, W.: Towards optimal binary code learning via ordinal embedding. In: AAAI. (2016) 1258–1265

[24] Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: CVPR. (2009) 112–119

[25] Yang, S., Luo, P., Loy, C.C., Tang, X.: From facial parts responses to face detection: A deep learning approach. In: ICCV. (2015) 3676–3684

[26] Zhang, J., Shan, S., Kan, M., Chen, X.: Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In: ECCV. (2014) 1–16

[27] Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NIPS. (2014) 1988–1996

[28] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: BMVC. (2015)

[29] Dong, Y., Liu, Y., Lian, S.: Automatic age estimation based on deep learning algorithm. Neurocomputing **187** (2016) 4–10

[30] Casado, I.H., Fernández, C., Segura, C., Hernando, J., Prati, A.: A deep analysis on age estimation. PR Letters **68** (2015) 239–249

[31] Kuang, Z., Huang, C., Zhang, W.: Deeply learned rich coding for cross-dataset facial age estimation. In: ICCVW. (2015) 338–343

[32] Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: CVPRW. (2015) 34–42

[33] Yang, H.F., Lin, B.Y., Chang, K.Y., Chen, C.S.: Automatic age estimation from face images via deep ranking. In: BMVC. (2015)

[34] Yang, X., Gao, B., Xing, C., Huo, Z., Wei, X., Zhou, Y., Wu, J., Geng, X.: Deep label distribution learning for apparent age estimation. In: ICCVW. (2015) 344–350

[35] Kleindessner, M., von Luxburg, U.: Uniqueness of ordinal embedding. In: COLT. (2014) 40–67

[36] Huang, K.H., Lin, H.T.: Cost-sensitive label embedding for multi-label classification. arXiv preprint arXiv:1603.09048 (2016)

[37] Weston, J., Bengio, S., Usunier, N.: WSABIE: scaling up to large vocabulary image annotation. In: IJCAI. (2011) 2764–2770

[38] Arias-Castro, E.: Some theory for ordinal embedding. arXiv preprint arXiv:1501.02861 (2015)

[39] Jr., K.R., Tesafaye, T.: MORPH: A longitudinal image database of normal adult age-progression. In: FG. (2006) 341–345

[40] Ebner, N.C., Riediger, M., Lindenberger, U.: Facesa database of facial expressions in young, middle-aged, and older women and men: Development and validation. Behavior Research Methods **42**(1) (2010) 351–362

[41] Escalera, S., Fabian, J., Pardo, P., Baró, X., Gonzàlez, J., Escalante, H.J., Misevic, D., Steiner, U., Guyon, I.: Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In: ICCVW. (2015) 243–251

[42] King, D.E.: Dlib-ml: A machine learning toolkit. JMLR **10** (2009) 1755–1758

[43] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. (2012) 1097–1105

[44] Chen, K., Gong, S., Xiang, T., Loy, C.C.: Cumulative attribute space for age and crowd density estimation. In: CVPR. (2013) 2467–2474

[45] Guo, G., Mu, G.: Human age estimation: What is the influence across race and gender? In: CVPR. (2010) 71–78

[46] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CVPR (2016)

[47] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)

[48] Yan, S., Wang, H., Tang, X., Huang, T.S.: Learning auto-structured regressor from uncertain nonnegative labels. In: ICCV. (2007) 1–8