Orthogonal Deep Features Decomposition for Age-Invariant Face Recognition

Yitong Wang, Dihong Gong, Zheng Zhou, Xing Ji, Hao Wang, Zhifeng Li(✉), Wei Liu(✉), and Tong Zhang

Tencent AI Lab, Beijing, China
{yitongwang,encorezhou,denisji,hawelwang,michaelzfli}@tencent.com,
gongdihong@gmail.com, wl2223@columbia.edu, tongzhang@tongzhang-ml.org

Abstract. As facial appearance is subject to significant intra-class variations caused by the aging process over time, age-invariant face recognition (AIFR) remains a major challenge in face recognition community. To reduce the intra-class discrepancy caused by the aging, in this paper we propose a novel approach (namely, Orthogonal Embedding CNNs, or OE-CNNs) to learn the age-invariant deep face features. Specifically, we decompose deep face features into two orthogonal components to represent age-related and identity-related features. As a result, identity-related features that are robust to aging are then used for AIFR. Besides, for complementing the existing cross-age datasets and advancing the research in this field, we construct a brand-new large-scale Cross-Age Face dataset (CAF). Extensive experiments conducted on the three public domain face aging datasets (MORPH Album 2, CACD-VS and FG-NET) have shown the effectiveness of the proposed approach and the value of the constructed CAF dataset on AIFR. Benchmarking our algorithm on one of the most popular general face recognition (GFR) dataset LFW additionally demonstrates the comparable generalization performance on GFR.

Keywords: Age-invariant face recognition
Convolutional neural networks · Cross-age face dataset

1 Introduction

As one of the most important topics in computer vision and pattern recognition, face recognition has attracted much attention from both academic and industry for decades [2,4,18,19,23,37,40,44]. With the evolution of deep learning, the performance of general face recognition (GFR) has been significantly improved in recent years, even higher than humans' abilities [24,32–35,39,43]. As a major challenge in face recognition, age-invariant face recognition (AIFR) is extremely valuable on various application scenarios, such as looking for lost children after decades, matching face images in different ages, etc. In contrast to GFR, AIFR involves more diversity with the significant intra-class variations caused by the

V. Ferrari et al. (Eds.): ECCV 2018, LNCS 11219, pp. 764–779, 2018.
https://doi.org/10.1007/978-3-030-01267-0_45

aging process and thus is more challenging. It is very often that the inter-class variation is much smaller than the intra-class variation in the presence of age variation, as illustrated in Fig. 1a. Figure 1b also exhibits the difficulty of AIFR where the same identity greatly varies in appearance with the aging process.

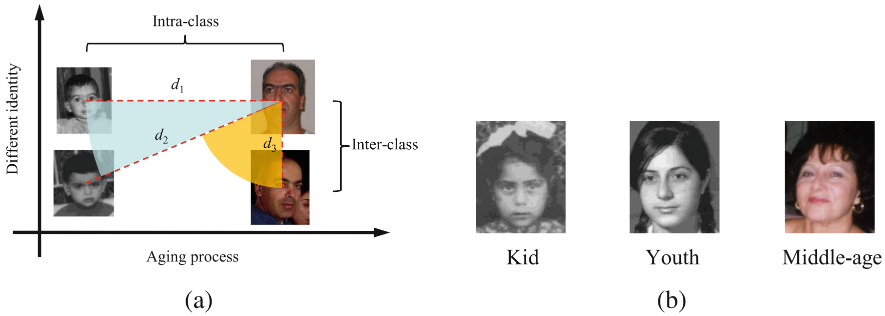


Fig. 1. The major challenge of AIFR: the intrinsic large intra-class variations in the aging process. (a) An example where intra-class distance is larger than inter-class distance. (b) The cross-age images for one subject in the FG-NET [1].

Recent AIFR researches primarily concentrate on two technical schemes: generative scheme and discriminative scheme. The generative scheme models the AIFR by synthesizing faces to one or more fixed age category then performs recognition with the artificial face representations [9,16,28]. Benefited from the advancement of the deep generative model, the generative scheme becomes more promising on AIFR as well [3,8,46]. However, the generative scheme still remains several significant shortcomings. Firstly, generative scheme usually separates the recognition process into two steps. Hence it is not easy for the generative models to optimize recognition performance in an end-to-end manner. Secondly, generation models are often unstable so the synthesizing face images will introduce additional noises, which may result in negative effects on the recognition process. Moreover, constructing an accurate, parametric generation model is fairly difficult since the aging process of humans' face is easily impacted by many latent factors such as social environments, diet, etc.

The discriminative scheme aims at constructing the sophisticated discriminative model to solve the problem of AIFR. Related works on discriminative model include [5–7,10,11,17,20–22]. By combining the deep learning algorithm, the discriminative scheme has achieved substantial improvement on AIFR. For example, Wen et al. [42] extended the HFA method [10] to a deep CNN model called latent factor guided convolutional neural networks (LF-CNNs), which achieved the state-of-the-art recognition accuracy in this field. Zheng et al. [47] also used the linear combination of jointly-learned deep features to represent identity and age information, which is similar to the HFA based deep CNN model.

In this paper, we aim at designing a new deep learning approach to effectively learn age-invariant components from features mixed with age-related information. The key idea of our approach is to decompose face features into age-related and identity-related components, where the identity-related component is age-invariant and suitable for AIFR. More specifically, inspired by a recent state-of-the-art deep learning GFR system with A-Softmax loss [24] where features of different identities are discriminated by different angles, we decompose face features in the spherical coordinate system which consists of radial coordinate r and angular coordinates $\phi_1, \ldots, \phi_n$. Then the identity-related components are represented with angular coordinates, and the age-related information is encoded with radial coordinate. Features separated by the two mutually orthogonal coordinate systems are then trained jointly with different supervision signals. Identity-related features are trained as a multi-class classification task supervised by identity labels with the A-Softmax loss, and age-related features are trained as a regression task supervised by age labels. As such, we extract age-invariant features from angular coordinates by separating age-related components with radial coordinates. Since face features are decomposed into mutually orthogonal coordinate systems, we name our approach as orthogonal embedding CNNs (OE-CNNs). A related work Decoupled Network also discussed how to decouple the CNN with orthogonal geometry in details. Nevertheless, this work merely studies the generalization of networks rather than specifically modeling the age into decomposed features in the AIFR application scenario. We verify the effectiveness of OE-CNNs with extensive experiments on three face aging datasets (MORPH Album2 [30], CACD-VS [5] and FG-NET [1]) and one GFR dataset (LFW [12]), and achieve the state-of-the-art performances.

The major contributions of this paper are summarized as follows:

1. We propose a new approach called OE-CNNs to tackle the problem on how to jointly model the age-related features and identity-related features in a deep CNN model. Based on the proposed model, age-invariant deep features can be effectively obtained for improved AIFR performance.
2. We introduce a new large-scale Cross-Age Face dataset, named CAF, to help advance the research in this field. This dataset contains more than 313,986 images from 4,668 identities. The face data in CAF has been manually cleaned in order to be noise-free.
3. We demonstrate the effectiveness of our proposed approach with several extensive experiments over three face aging datasets (MORPH Album2 [30], CACD-VS [5] and FG-NET [1]) and one GFR dataset (LFW [12]). The experimental results have shown the superior performance of the proposed approach over the state-of-the-art either on AIFR or GFR.

2 Proposed Approach

2.1 Orthogonal Deep Features Decomposition

Two certain difficulties involved in AIFR include the considerable variations of the identical individual in different age categories (intra-class variations) caused by aging process (such as shape changes, texture changes, etc.), and

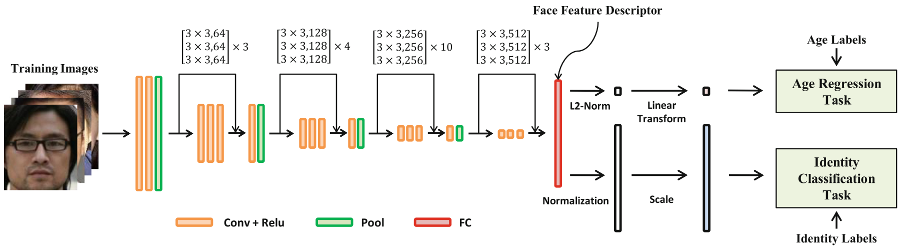


Fig. 2. The proposed ResNet-Like CNN architecture.

the inevitable mixture of unrelated components in the deep features extracted from a general deep CNN model. Large intra-class variation usually leads to erroneous identification on a pair of faces from the same individual at different ages. The mixed features (age features and identity features) potentially reduce the robustness of recognizing cross-age faces. To address this, we propose a new approach called orthogonal embedding CNNs. Below we first walk through the problem of deep AIFR in detail.

Given an observed Fully-Connected (FC) feature x extracted from the deep CNN model, we decompose it into two components (vectors). One is identity-related component x_{id} and the other is age-related component x_{age}. Thus, after removing x_{age} from x, we can obtain x_{id} that is supposed to be age-invariant. Recent works [10,42,47] use a linear combination to model x_{age} and x_{id} as the solution. In this paper, we propose a new approach to model x_{age} and x_{id} in an orthogonal manner with deep convolutional neural networks. Inspired by A-Softmax [24], where features of different identities are discriminated by different angles, we decompose feature x in spherical coordinate system $x_{sphere} = \{r; \phi_1, \phi_2, ..., \phi_n\}$. The angular components $\{\phi_1, \phi_2, ..., \phi_n\}$ represent identity-related information, and the rest radial component r is used to encode age-related information. Formally, $x \in R^n$ is decomposed under x_{sphere} as

$$x = x_{age} \cdot x_{id}, \tag{1}$$

where $x_{age} = ||x||_2$, and $x_{id} = \{\frac{x_1}{||x||_2}, \frac{x_2}{||x||_2}, ..., \frac{x_n}{||x||_2}\}$, with $||x_{id}||_2 = 1$. Here $||.||_2$ represents for L_2 norm, and x_n is the n-th component of x. For convenience, we will use n_x to represent for $||x||_2$ and $\tilde{x}$ for $\frac{x}{||x||_2}$.

2.2 Multi-task Learning

According to Eq. 1, feature x output from the last FC layer is decomposed into x_{age} and x_{id}. In this part, we describe a multi-task based learning algorithm to jointly learn these features. An overview of the proposed CNN model is illustrated in Fig. 2.

Learning Age-Related Component. In order to dig out the intrinsic clues of age information, we utilize an age estimation task to learn the relationship between the component x_{age} (n_x) and the ground truth of age. For simplicity, linear regression is adopted to the age estimation task, and the regression loss can be formulated as follows:

$$L_{age} = \frac{1}{2M} \sum_{i=1}^{M} ||f(n_{x_i}) - z_i||_2^2 \tag{2}$$

where n_{x_i} is the L_2 norm of the i-th embedding feature x_i, z_i is the corresponding i-th age label. $f(x)$ is a mapping function aimed to associate n_{x_i} and z_i. Since the L_2 norm n_{x_i} is a scalar, we use linear polynomial $f(x) = k \cdot x + b$ as the mapping function. We also explored other more complicated functions such as non-linear multi-layer perceptron network, but they did not perform as well as a simple linear transformation. We believe this is because a more complicated model overfits the underlying feature which is one-dimensional here.

Learning Identity-Related Component. When performing face verification or identification, $\tilde{x}$ is the only part which participates in the final similarity measure. Thus, the identity-related component x_{id} should be as discriminative as possible. Following the recent state-of-the-art GFR algorithm A-Softmax [24], we use a similar loss function to increase classification margin between different training persons in angular space:

$$L_{id} = \frac{1}{M} \sum_{i=1}^{M} - \log\left(\frac{e^{s \cdot \psi(\theta_{y_i,i})}}{e^{s \cdot \psi(\theta_{y_i,i})} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_{j,i})}}\right) \tag{3}$$

in which $\psi(.)$ is defined as $\psi(\theta_{y_i,i}) = (-1)^k \cos(m\theta_{y_i,i}) - 2k$, $\theta_{y_i,i}$ is the angle between the i-th feature $\tilde{x}_i$ and label y_i's weight vector, $\theta_{y_i,i} \in [\frac{k\pi}{m}, \frac{(k+1)\pi}{m}]$, and $k \in [0, m-1]$. $m \geq 1$ is an integer hyper-parameter that controls the size of angular margin, and $s > 0$ is an adjustable scale factor introduced to compensate the learning of Softmax. From the geometric perspective, Eq. 3 adds a constraint which guarantees the angle of the feature x with its corresponding weight vector should less than $\frac{1}{m}$ of the angle between the feature x and any other weight vectors. Consequently, the margin between two arbitrary classes can be increased. Compared with the original A-Softmax, Eq. 3 replaces L_2 norm of $\tilde{x}$ with an adjustable scalar factor s. In our model, according to Eq. 1, $||\tilde{x}||_2$ is always equal to 1. Thus, it is necessary to introduce an extra free variable to compensate for the loss of L_2 norm.

Overall, the two losses are combined to a multi-task loss for jointly optimizing, as below:

$$L = L_{id} + \lambda L_{age} \tag{4}$$

where λ is a scalar hyper-parameter to balance the two losses. Equation 4 is used to guide the learning of our CNN model in the training phase. In the testing phase, only the identity-related component x_{id} is used for the AIFR task.

2.3 Discussion

Compared with HFA Based AIFR Methods. The HFA based AIFR methods [10,11,42] suggest modeling the identity-related component and age-related

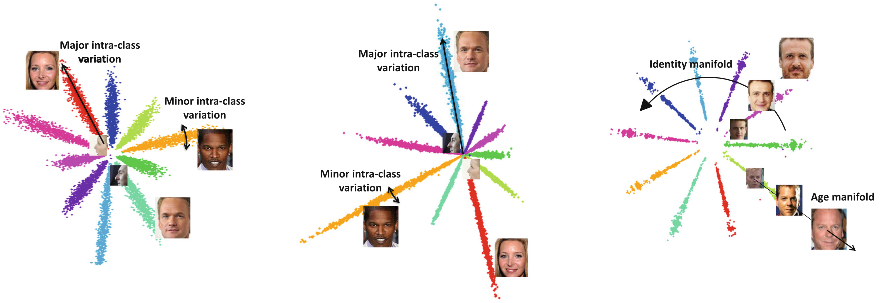
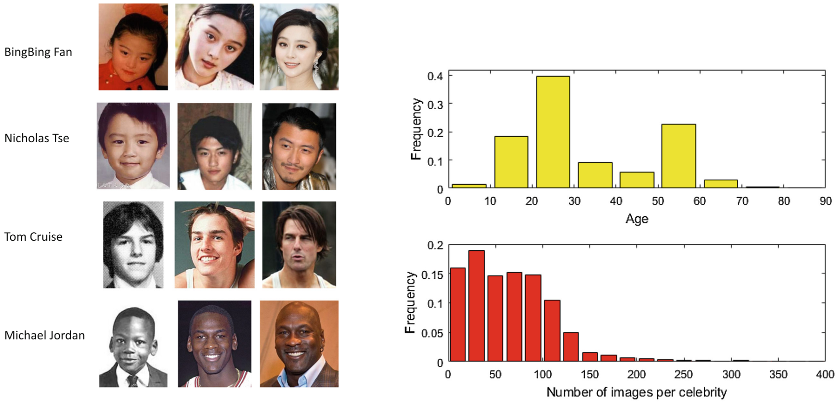


Fig. 3. Visualization of deep features learned with Softmax (Left), A-Softmax (Middle) and the proposed algorithm (Right). It is noteworthy that only 10 individuals are used to train CNN models, and the output dimension is set to 2. Colors are used to distinguish identities, and placement of face images is based on the corresponding features.

component of features by the simple linear combination. Specifically, given a feature x, the HFA based methods decompose the x as $x = m + Ux_{age} + Vx_{id} + \varepsilon$, where m is the mean feature regarding identity-related component, ε is the additional noise and U, V are the transformation matrices for identity-related component x_{id} and age-related component x_{age} respectively. The major advancements of the proposed approach over the HFA based methods are described in the following aspects: Firstly, the proposed approach revises the decomposition of x in the HFA based methods to the multiplication of hidden components x_{id} and x_{age}, which is more intuitive and concise to model the unrelated components with less extra hyper-parameters. Secondly, we explicitly project the identity features on a hypersphere to match the cosine similarity measurement for effectively combining the improvement strategies based on the Softmax loss and the margin of decision boundaries. Thirdly, the HFA based methods have to iteratively run the EM algorithm in contrast to our approach which jointly trains the network in the desirable end-to-end manner of feature learning. For the foregoing reasons, our method is more recommendable to be embedded into CNN framework for the purpose of learning age-invariant features, as supported by our experimental results.

Compared with SphereFace. SphereFace [24] introduces A-Softmax loss to learn the angular margin between identities for GFR. Though we train the identity-related component with a loss function similar to A-Softmax, the proposed algorithm takes advantage of the age information to explicitly train age-related component with an additional age regression task (Eq. 2). To intuitively investigate the impact by introducing such additional age regression task, we construct a toy example to compare features learned by Softmax, A-Softmax and our proposed algorithm. Specifically, we train CNN models with 10 individuals and set the output dimension of feature x as 2. For simplicity we let $f(x) = x$ (see Eq. 2) in this case. Figure 3 is the visualization for training



(a) (b)

Fig. 4. Overview of the CAF dataset. (a) Example images of CAF. Note that since our images are collected from Internet, CAF not only varies in ages but also in poses, races, etc. (b) The distribution of CAF. Top: The distribution of the number of different ages. Bottom: The distribution of the number of different identities.

features. Based on this example, we conclude that: (1) features of different persons are discriminated mostly by angles, which intuitively justifies our decomposition design; (2) both A-Softmax and the proposed algorithm have noticeably larger classification margins than Softmax, as a result of the A-Softmax loss; (3) most importantly, for our model age of a person is reflected in radial direction (e.g. larger L_2 norms for older faces), while the other two models do not have this property. We believe this property further constrains the training problem, which reduces the risk of over-fitting and consequently leads to superior performance for AIFR.

Generalization of Our Approach. One of the noticeable highlights of the proposed algorithm is its generalization capability. Intuitively, our method is specifically designed to fit cross-age training data. However, the experimental results surprisingly unfold the excellent performance of the proposed method even trained with general training data (as shown in Sect. 4.4). Furthermore, as the objective of the algorithm is to generate identity-related features, the proposed algorithm is not only suitable for AIFR but also for GFR. Finally, The age component can be easily generalized to any other common component such as pose, illumination, emotion, etc.

3 Large-Scale Cross-Age Face Dataset (CAF)

In order to further motivate the development of AIFR and enrich the capability of the current model, a dataset with a large age gap is urgently needed. Besides, the dataset size should be large enough to avoid overfitting. To this end, we collect a new dataset with a large number of cross-age celebrities' faces, named large-scale Cross-Age Face dataset (CAF).

3.1 Dataset Collection

To build the cross-age celebrity dataset, it is inevitable to collect celebrity's name to form a list. The collected names in the list come from multiple sources such as IMDB, Forbes celebrity list, child actors name list from Wikipedia, etc. This guarantees the comparatively large age gap in the later data collection. Next, we iteratively search the name in the list by the Google Search Engine. Each searching term has been thresholded to a certain number, that is, we keep the name in the list if the number of responses exceeds a certain threshold, which ensures the sufficient number of data for each celebrity. Moreover, to the best of our knowledge, the current public cross-age datasets have very limited Asian individuals. For the purpose of increasing the diversity of our cross-age dataset, we collect a large number of Asian celebrities. After filtering the name list, we download the face images on several commercial image search engine (such as Google, Baidu) querying by the celebrity's name companied with several keywords like *yearbook*, *past and now*, *childhood*, *young*, *from young to old*, etc., to obtain the face images with different age categories. The data cleaning is performed thereafter. Specifically, we apply face detection algorithm MTCNN [14] to filter the images without any faces, then manually wipe off the near-duplicates and false face images (faces do not belong to that celebrity). Finally, we delete some of the images that have a large proportion in a certain age category to keep the age distribution more balanced.

Table 1. Comparison over cross-age datasets.

Dataset	CAF	IMDB-WIKI [31]	CACD [5]	MORPH [30]	AgeDB [26]	FG-NET [1]
# Images	313K	523K	163K	78K	16K	1K
# Subjects	4,668	20,284	2,000	20,000	568	82
Noise-free	Yes	No	Yes	Yes	Yes	Yes

3.2 Dataset Statistics

Following the above labeling and cleaning process, we construct a cross-age face dataset which totally includes about 313,986 face images from 4,668 identities. Each identity has approximately 80 face images. All of these images have been carefully and manually annotated. Example images of the dataset are shown in Fig. 4a. Considering the lack of exact age information, we utilize the public pre-trained age estimation model DEX [31] to predict the rough age label for each face image. Figure 4b shows the distribution histogram of CAF. One can observe our data are well-distributed in every possible age category. Table 1 fairly compares our dataset with existing released cross-age datasets. It is clear that except IMDB-WIKI [31], we have the comparatively largest scale in terms of the number of pictures and the number of individuals. Furthermore, as IMDB-WIKI is collected by automatically online crawling, some of the downloaded data might be redundant and noise-severe. Superior to IMDB-WIKI, CAF has minimized the noise data by manually annotating.

4 Experiments

For a direct and fair comparison to the existing work in this field, we evaluate our approach on existing public-domain cross-age face benchmark datasets MORPH Album 2 [30], CACD-VS [5] and FG-NET [1]. We also evaluate our algorithm on LFW [12] for verifying the generalization performance on GFR.

4.1 Implementation Details

The training set is composed of two parts: a cross-age face dataset and a general face dataset (without cross-age face data). The cross-age face dataset that we use is the collected CAF dataset introduced in Sect. 3 while the general face dataset consists of three public face datasets: CASIA-WebFace [45], VGG Face [29] and celebrity+ [25]. The same identities appeared in different datasets are carefully merged together. Since our testing dataset contains MORPH, CACD-VS, FG-NET, and LFW, we have excluded these data from the training set. Finally, our training set contains 1,765,828 images with 19,976 identities in total, which includes 313,986 cross-age face images with 4,668 identities and 1,451,842 general face images with 17609 identities respectively. In addition, the age label predicted from the public pre-trained age estimation model DEX [31] is treated as the regression target of Euclidean loss. Prior to training stage, we perform the same pre-processing on both training set and testing set: Using MTCNN [14] to detect the face and facial key points in images, then applying similarity transformation to crop the face patch to 112 × 96 pixels according to the 5 facial key points (two eyes, nose and two mouth corners), finally normalizing the cropped face patch by subtracting 127.5 then divided by 128. The proposed loss in Eq. 3 serves as the supervisory signal of identity classification. In terms of the age branch, we use Euclidean loss function to guide the network to learn the age label. The hyper-parameters m, s mentioned in Eqs. 3 and 4 are set to 4, 32 according to the recommendations of [24,38]. For the factor λ, we empirically selected an optimal value 0.01 to balance the two losses. All models are trained with Caffe [13] framework and optimized with stochastic gradient descent (SGD) algorithm. Training batch size is set to 512 and the number of iterations is set to 21 epochs. The initial learning rate is set to 0.05 and the training process adaptively decreases the learning rate 3 times when the loss becomes stable (roughly at the 9-th, 15-th and 18-th epoch).

4.2 Experiments on the MORPH Album 2 Dataset

Following [10,11,17,42], in this study we use an extended version of MORPH Album 2 dataset [30] for performance evaluation. It has 78,000 face images of

20,000 identities in total. The data has been split into training and testing set. The training set contains 10,000 identities. The rest of 10,000 identities belong to testing set where each identity has 2 photos with a large age gap. The testing data have been divided into gallery set and probe set. We follow the testing procedure given by [10] to evaluate the performance of our algorithm. We set up several schemes for comparison including: (1) **Softmax:** the CNN-baseline model trained by the original Softmax loss, (2) **A-Softmax:** the CNN-baseline model guided by the A-Softmax loss, (3) **OE-CNNs:** the proposed approach, and (4) other recently proposed top-performing AIFR algorithm in the literatures.

Firstly, we compare the proposed approach to baseline algorithms that are most related to the proposed algorithm to demonstrate its effectiveness. Table 2 compares the rank 1 identification rates testing on 10,000 subjects of Morph Album 2 over Softmax, A-Softmax, and OE-CNNs, with and without CAF dataset. As shown in the table, The proposed **OE-CNNs** significantly outperforms both Softmax and A-Softmax under both settings. Specifically, though we've used similar loss function with A-Softmax for training the identity-related features, **OE-CNNs** noticeably improves the performance of A-Softmax, which confirms the effectiveness of our features decomposition method for AIFR. Note that, all compared networks have the same base network (from input to FC layer). When comparing performances trained with and without CAF dataset, we can see that with CAF the identification rate improves consistently for all systems, which confirms that the CAF dataset is valuable to AIFR research.

Secondly, for ensuring a fair comparison with other methods, we neglect the CAF dataset and conduct an experiment with the same training data as related work [42] has used. Specifically, WebFace [45], celebrity+ [25] and CACD [5] form the training set to train a CNN base model. The trained model is later fine-tuned with Morph training data. Table 3 depicts our result compared with other methods. There are conventionally two evaluation schemes on Morph benchmark: testing on 10,000 subjects or 3,000 subjects. For fairly comparing against other methods, we evaluate the proposed OE-CNN approach on both schemes. As can be seen in Table 3, the OE-CNN approach shows its capability by substantially outperforming all other methods in both two evaluation schemes. Particularly, our method surpasses the LF-CNN model by 1.0% and AE-CNN model by 0.5%, which is an outstanding improvement on the accuracy level above 98%.

4.3 Experiments on the CACD-VS Dataset

CACD dataset comprises comprehensively 163,446 images from 2,000 distinct celebrities. The age ranges from 10 to 62 years old. This dataset collects the celebrity's images with the effect of various illumination condition, different poses and makeup, which can effectively reflect the robustness of the AIFR algorithm. CACD-VS is a subset of CACD which is picked from CACD to composes 2,000 pairs of positive sample and 2,000 pairs of negative samples, and 4,000 pairs of samples in total. We follow the pipeline of [5] to calculate the similarity score of all sample pairs and the ROC curves and its corresponding AUC. We take 9 folds from 10 folds that have already been separated officially to compute threshold

Table 2. Performance comparisons of different baselines on Morph Album 2.

Training dataset	Method	Rank-1 identification rates
Public datasets	Softmax	94.84%
Public datasets	A-Softmax	96.27%
Public datasets	**OE-CNNs**	**97.46%**
Public datasets + CAF	Softmax	95.49%
Public datasets + CAF	A-Softmax	96.59%
Public datasets + CAF	**OE-CNNs**	**98.57%**

Table 3. Performance comparisons of different approaches on Morph Album 2.

Method	#Test subjects	Rank-1 identification rates
HFA [10]	10,000	91.14%
CARC [5]	10,000	92.80%
MEFA [11]	10,000	93.80%
MEFA+SIFT+MLBP [11]	10,000	94.59%
LPS+HFA [17]	10,000	94.87%
LF-CNNs [42]	10,000	97.51%
OE-CNNs	10,000	**98.55%**
GSM [21]	3,000	94.40%
AE-CNNs [47]	3,000	98.13%
OE-CNNs	3,000	**98.67%**

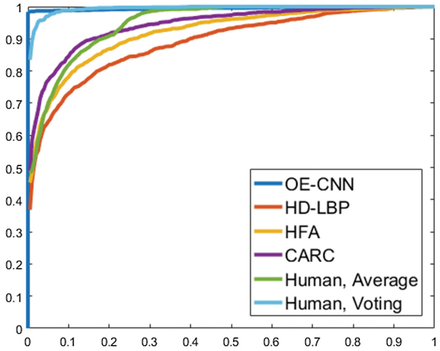


Fig. 5. ROC comparisons of different approaches on CACD-VS.

references and use this threshold to evaluate on the rest of 1 fold. By repeating this procedure 10 times, we finally calculate the average accuracy as another measure.

Table 4. Performance comparisons of different approaches on CACD-VS.

Method	Acc	AUC
High-dimensional LBP [7]	81.6%	88.8%
HFA [10]	84.4%	91.7%
CARC [5]	87.6%	94.2%
LF-CNNs [42]	98.5%	99.3%
Human, Average [6]	85.7%	94.6%
Human, Voting [6]	94.2%	99.0%
Softmax	98.4%	99.4%
A-Softmax	98.7%	99.5%
OE-CNNs	**99.2%**	**99.5%**

The results of all the baselines are shown in Table 4 and Fig. 5. As illustrated, the proposed OE-CNN approach significantly outperforms all the other baselines. Furthermore, our approach also surpasses the human-level performance, which demonstrates the effectiveness of our proposed age-invariant deep features.

4.4 Experiments on the FG-NET Dataset

The FG-NET dataset consists of 1,002 pictures from 82 different identities, each identity has multiple face images with huge variability in the age covering from child to elder. Following the evaluation protocols of Megaface challenge 1 (MF1) [15] and Megaface challenge 2 (MF2) [27] we employ the 1 million images from Flickr as the distractor set. Particularly, under the small protocol of MF1, we reduce our training data to 0.5 million images from 12,073 identities in the training phase. The cross-age face images in FG-NET servers as the probe set in which a probe image is compared against each image from distractor set. We evaluate the rank-1 performance of the presented algorithm under the protocols of MF1 and MF2, as shown in Tables 5 and 6, respectively.

Under the small protocol of MF1, the proposed method not only obtains a significant performance improvement over Softmax and A-Softmax baseline but also surpasses the existing methods (including a specific age-invariant method TNVP [8]) by a clear margin. Under the protocol of MF2, all the algorithms need to be trained using the same training dataset (which does not involve the cross-age training data) provided by MF2 organizer. It is encouraging to see that our algorithm also outperforms all other methods with a large margin, which strongly proves the effectiveness of our algorithm on AIFR.

4.5 Experiments on the LFW Dataset

LFW is a very famous benchmark for general face recognition. The dataset has 13,233 face images from 5,749 subjects acquiring from the arbitrary environment.

Table 5. Performance comparisons of different approaches under the protocols of MF1 [15] on FG-NET.

Method	Protocol	Rank-1 identification rates
FUDAN-CS_SDS [41]	Small	25.56%
SphereFace [24]	Small	47.55%
TNVP [8]	Small	47.72%
Softmax	Small	35.11%
A-Softmax	Small	46.77%
OE-CNNs (single-patch)	Small	52.67%
OE-CNNs (3-patch ensemble)	Small	**58.21%**

Table 6. Performance comparisons of different approaches under the protocol of MF2 [27] on FG-NET.

Method	Protocol	Rank-1 identification rates
GRCCV	Large	21.04%
NEC	Large	29.29%
3DiVi	Large	35.79%
GT-CMU-SYSU	Large	38.21%
OE-CNNs (single-patch)	Large	**53.26%**

Table 7. Performance comparisons of different approaches on LFW.

Method		Images	Networks	Acc
General approaches	DeepFace [36]	4M	3	97.35%
	FaceNet [32]	200M	1	99.65%
	DeepID2+ [35]	–	25	99.47%
	Center loss [43]	0.7M	1	99.28%
	SphereFace [24]	0.5M	1	99.42%
Cross-age approaches	LF-CNNs [42]	0.7M	1	99.10%
	OE-CNNs	0.5M	1	**99.35%**
	OE-CNNs	1.7M	1	**99.47%**

We experiment our algorithm on LFW following the official unrestricted with labeled outside data protocol. We test our model on 6,000 face pairs. The training data are disjoint from the testing data. Table 7 exhibits our results. One can see that the proposed OE-CNN approach achieves comparable performance without any ensemble trick to the state-of-the-art approaches, which demonstrates the excellent generalization ability of the proposed approach. Additionally, after we expand the training dataset to 1.7M (including CAF dataset), the performance of OE-CNNs further improves to 99.47%, which also proves that our CAF dataset is not only valuable for AIFR but also helpful for GFR.

5 Conclusion

AIFR is a remained challenging computer vision task on account of the aging process of the human. Inspired by pioneering work and the observation of hidden components, this paper proposes a novel approach which separates deep face feature into the orthogonal age-related component and identity-related component to improve AIFR. The highly discriminative age-invariant features can be consequently extracted from a multi-task deep CNN model based on the proposed approach. Furthermore, we build a large cross-age celebrity dataset named CAF that is both noise-free and vast in the number of images. As a part of training data, CAF greatly boosts the performance of the models for AIFR. Extensive evaluations of several face aging datasets have been done to show the effectiveness of our orthogonal embedding CNN (OE-CNN) approach. More studies on how to incorporate the generative scheme and improve the discriminative scheme will be explored in our future work to benefit the AIFR community.

References

1. FG-NET Aging Database. http://www.fgnet.rsunit.com/
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) (2006)
3. Antipov, G., Baccouche, M., Dugelay, J.L.: Face Aging With Conditional Generative Adversarial Networks. In: IEEE International Conference on Image Processing (ICIP) (2017)
4. Belhumeur, P., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) (1997)
5. Chen, B.C., Chen, C.S., Hsu, W.H.: Cross-age reference coding for age-invariant face recognition and retrieval. In: European Conference on Computer Vision (ECCV) (2014)
6. Chen, B.C., Chen, C.S., Hsu, W.H.: Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. IEEE Trans. Multimed. **17**(6), 804–815 (2015)
7. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3025–3032 (2013)
8. Duong, C.N., Quach, K.G., Luu, K., Savvides, M., et al.: Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
9. Geng, X., Zhou, Z.H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) (2007)
10. Gong, D., Li, Z., Lin, D., Liu, J., Tang, X.: Hidden factor analysis for age invariant face recognition. In: International Conference on Computer Vision (ICCV) (2013)

11. Gong, D., Li, Z., Tao, D., Liu, J., Li, X.: A maximum entropy feature descriptor for age invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5289–5297 (2015)
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst (2007)
13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 2016 ACM on Multimedia Conference (ACM MM) (2014)
14. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multi-task cascaded convolutional networks. Signal Process. Lett. **23**(10), 1499–1503 (2016)
15. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
16. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) (2002)
17. Li, Z., Gong, D., Li, X., Tao, D.: Aging face recognition: a hierarchical learning model based on local patterns selection. IEEE Trans. Image Process. (TIP) **25**(5), 2146–2154 (2016)
18. Li, Z., Lin, D., Tang, X.: Nonparametric discriminant analysis for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI) **31**, 755–761 (2009)
19. Li, Z., Liu, W., Lin, D., Tang, X.: Nonparametric subspace analysis for face recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2005)
20. Li, Z., Park, U., Jain, A.K.: A discriminative model for age invariant face recognition. IEEE Trans. Inf. Forensics Secur. (TIFS) (2011)
21. Lin, L., Wang, G., Zuo, W., Feng, X., Zhang, L.: Cross-domain visual matching via generalized similarity measure and feature learning. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) **39**(6), 1089–1102 (2017)
22. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: Face verification across age progression using discriminative methods. IEEE Trans. Inf. Forensics Secur. (TIFS) (2010)
23. Liu, W., Li, Z., Tang, X.: Spatio-temporal embedding for statistical face recognition from video. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) European Conference on Computer Vision (ECCV) (2006)
24. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: SphereFace: deep hypersphere embedding for face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
25. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: International Conference on Computer Vision (ICCV) (2015)
26. Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S.: Agedb: the first manually collected in-the-wild age database. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
27. Nech, A., Kemelmacher-Shlizerman, I.: Level playing field for million scale face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
28. Park, U., Tong, Y., Jain, A.K.: Age-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) (2010)
29. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (BMVC) (2015)

30. Ricanek, K., Tesafaye, T.: Morph: a longitudinal image database of normal adult age-progression. In: International Conference on Automatic Face and Gesture Recognition (2006)
31. Rothe, R., Timofte, R., Gool, L.V.: Dex: deep expectation of apparent age from a single image. In: International Conference on Computer Vision Workshops (ICCVW), December 2015
32. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: a unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
33. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: Advances in Neural Information Processing Systems (NIPS) (2014)
34. Sun, Y., Liang, D., Wang, X., Tang, X.: Deepid3: face recognition with very deep neural networks. arXiv preprint arXiv:1502.00873 (2015)
35. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
36. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
37. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Conference on Computer Vision and Pattern Recognition (CVPR) (1991)
38. Wang, F., Xiang, X., Cheng, J., Yuille, A.L.: NormFace: L_2 hypersphere embedding for face verification. In: Proceedings of the 2017 ACM on Multimedia Conference (ACM MM) (2017)
39. Wang, H., Wang, Y., Zhou, Z., Ji, X., Li, Z., Gong, D., Zhou, J., Liu, W.: Cosface: large margin cosine loss for deep face recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
40. Wang, X., Tang, X.: A unified framework for subspace face recognition. IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) (2004)
41. Wang, Z., He, K., Fu, Y., Feng, R., Jiang, Y.G., Xue, X.: Multi-task deep neural network for joint face recognition and facial attribute prediction. In: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR) (2017)
42. Wen, Y., Li, Z., Qiao, Y.: Latent factor guided convolutional neural networks for age-invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
43. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision (ECCV), pp. 499–515 (2016)
44. Xiong, Y., Liu, W., Zhao, D., Tang, X.: Face recognition via archetype hull ranking. In: International Conference on Computer Vision (ICCV) (2013)
45. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
46. Zhang, Z., Song, Y., Qi, H.: Age progression/regression by conditional adversarial autoencoder. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
47. Zheng, T., Deng, W., Hu, J.: Age estimation guided convolutional neural network for age-invariant face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)