Neurocomputing 222 (2017) 62–71

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Age invariant face recognition and retrieval by coupled auto-encoder networks

Chenfei Xu, Qihe Liu, Mao Ye*

School of Computer Science and Engineering, Center for Robotics, University of Electronic Science and Technology of China, Chengdu 611731, PR China

ARTICLE INFO

Communicated by Dr Jiwen Lu

Keywords:
Face recognition
Age invariant
Auto-encoder

ABSTRACT

Recently many promising results have been shown on face recognition related problems. However, age-invariant face recognition and retrieval remains a challenge. Inspired by the observation that age variation is a nonlinear but smooth transform and the ability of auto-encoder network to learn latent representations from inputs, in this paper, we propose a new neural network model called coupled auto-encoder networks (CAN) to handle age-invariant face recognition and retrieval problem. CAN is a couple of two auto-encoders which bridged by two shallow neural networks used to fit complex nonlinear aging and de-aging process. We further propose a nonlinear factor analysis method to nonlinearly decompose one given face image into three components which are identity feature, age feature and noise, where identity feature is age-invariant and can be used for face recognition and retrieval. Experiments on three public available face aging datasets: FGNET, CACD and CACD-VS show the effectiveness of the proposed approach.

1. Introduction

Age-invariant face recognition and retrieval is a challenging problem on face recognition research because one person can exhibit substantially different appearance at different ages which significantly increase the recognition difficulty. And it is becoming increasingly important and has a wide application, such as finding missing children, identifying criminals and passport verification. A traditional method proposed in [1,2] is to synthesis a face image to match the image at a target age before recognition. They try to construct a 2D/3D model to compensate for the age variation degrading the face recognition performance. However, these generative models strongly depend on parameters assumptions, accurate age labels and relatively clean training data, so they do not work well in real-world face recognition.

To address this problem, some discriminative methods [3–7] are proposed. Most of these methods attempt to design an appropriate feature representation and an effective matching framework. Typically, Li et al. [6] combined scale invariant feature transform (SIFT) [8] and multi-scale local binary pattern (MLBP) [9] as local feature representations for recognition but this method does not consider age information. Recently, [4,5] proposed an approach based on factor analysis. It considers the face image feature of one person can be expressed as combination of an identity-specific component and an age-related component. In the test phase, this method computes matching score of a given pair of images based on identity component (age-invariant). However, these methods are all linear models that their expressive power is limited and need a complex inference.

Motivated by the ability of auto-encoder to learn latent representations from inputs and the observation that age variation is a nonlinear but smooth transform, we propose a new neural network model called coupled auto-encoder networks (CAN). Given a pair of images of one person, we first choose two auto-encoders to accept these two images respectively as inputs to reconstruct them. Then, we leverage two shallow neural networks as a *bridge* to connect these two auto-encoders. We fit aging and de-aging process by these shallow neural networks due to the fact that any one single-hidden-layer neural network can fit any complex smooth function [10]. Further, a nonlinear factor analysis method is applied to the hidden layers of CAN, in which the representation of a face image is decomposed into three components: identity feature which is age-invariant, age feature which is identity-independent and noise. In the end, we apply PCA and LDA method [11] on the identity feature to form a more compressed and discriminative feature as the final age-invariant representation for face recognition and retrieval.

Our main contributions are: (1) a new model of age invariant face recognition and retrieval is proposed based on a couple of auto-encoder networks. Our approach is evaluated on three public face aging dataset, FGNET [12], CACD [13] and CACD-VS [14]; (2) we propose a nonlinear factor analysis method to separate identity feature from face representation. Compared with the similar methods based on linear

* Corresponding author.
E-mail address: cvlab.uestc@gmail.com (M. Ye).

http://dx.doi.org/10.1016/j.neucom.2016.10.010
Received 21 June 2016; Received in revised form 3 October 2016; Accepted 8 October 2016
Available online 18 October 2016

factor analysis proposed in [4,5], our method can obtain better identity feature.

The rest of this paper is organized as follows. Section 2 discusses related works. Section 3 describes the proposed approaches and details the coupled auto-encoder networks (CAN). Section 4 provides the experimental results. Section 5 concludes the paper.

2. Related works

Most existing works on age-related face analysis problems focus on age estimation [15–26] and age simulation [27–30,1,2]. The works on age-invariant recognition are limited and traditional methods fall into two categories. Generative methods proposed in [1,2] try to construct a 2D/3D face aging pattern space to synthesis a face image to match the target face image before recognition. However, these methods strongly depend on parameters assumptions, accurate age labels and relatively clean training data, so they do not work well in real-world face recognition.

Recently, some discriminative methods [3,31,4–7] are proposed and get good results. Ling et al. [7] use gradient oriented pyramid with support vector machine for face verification. Li et al. [6] design a densely sampled local feature description scheme combining scale invariant feature transform and multi-scale local binary pattern to improve face matching accuracy. Gong et al. [4] propose hidden factor analysis that tries to separate age variation from person-specific features for face recognition, and further propose a maximum entropy feature descriptor with identity factor analysis in [5] to improve this method. Lu et al. [31] propose a compact binary face descriptor for face representation and recognition. In [3], a new feature descriptor called local pattern selection (LPS) is proposed for aging face recognition.

Data-driven methods based on a reference set also have been used to improve age-invariant face recognition and retrieval. The authors in [13,14] propose a coding framework called Cross-Age Reference Coding (CARC) using CACD [13], a new large-scale face aging dataset, as a reference set to encode the low features of a face image into age-invariant representations.

Some deep learning models [32–39] also have been proposed. Wen et al. [32] propose a deep face recognition framework called latent factor guided convolutional neural network (LF-CNN) to significantly improve age invariant face recognition performance. With a model called latent identity analysis (LIA), they extract the age invariant features. Similarly, [33–35] respectively propose different convolutional neural network architecture to address the age invariant face recognition problem. Ref. [36] presents a generalized similarity model (GSM) and integrate it with the feature representation learning via deep convolutional neural networks for age-invariant face recognition. In [37], a deep aging face verification (DAFV) architecture is proposed including two modules called aging pattern synthesis module and aging face verification module. Ref. [39] combines deep convolutional neural networks with local binary pattern histograms (DCNN+LBPH) for face verification across aging. Ref. [38] presents a new joint feature learning (JFL) approach and stacks this model into a deep architecture to exploit hierarchical information for face representation.

Auto-encoder attempts to learn hidden representations automatically from inputs. It has been successfully applied in many computer vision problems. As a typical unsupervised learning method, auto-encoder [40,41] has shown its efficiency in many face-related recognition problems. Kan et al. [42] propose a stacked progressive auto-encoder for face recognition across poses. Liu et al. [43] use a sparse auto-encoder for facial expression recognition and Zhang et al. [44] propose an iterative stacked de-noising auto-encoder to recognize faces with partial occlusions. Liu et al. [37] use deep aging-aware de-noising auto-encoder for aging pattern synthesis (Fig. 1).

3. Proposed approaches

In this section, we describe the proposed approaches. We first overview the CAN model and then detail it. Next we present our training algorithm followed by the face matching method.

3.1. Overview

An overview of CAN is shown in Fig. 2. Structurally CAN is composed of two identical auto-encoders and two single-hidden-layer neural networks as a bi-directional *bridge*.

Inputs of CAN are training facial image pairs of different persons denoted as $T = \{\mathbf{x}_1^i, \mathbf{x}_2^i\}(\mathbf{x}_1^i, \mathbf{x}_2^i \in \mathbf{R}^n, i = 1, 2, 3, \ldots, N)$ where N is the total number of training image pairs. For one person, our goal is to encode age-invariant feature from inputs for recognition and retrieval, and a nonlinear factor analysis model is given as:

$$\mathbf{x} = \sigma(\mathbf{I}, \mathbf{A}, \xi), \tag{1}$$

where $\mathbf{x}$ represents inputs and $\sigma(\cdot)$ is a nonlinear function defined by CAN. The equation above means that a facial image can be decomposed into three components nonlinearly, i.e., $\mathbf{I}$ represents identity feature which is age-invariant, $\mathbf{A}$ represents age feature which is identity-independent and $\boldsymbol{\xi}$ represents noise which could be any factors deviate from our model.

Concretely, as shown in Fig. 2, $\mathbf{x}_1$ and $\mathbf{x}_2$ represent the younger and older facial image inputs of the same person. $\mathbf{I}_j$, $\mathbf{A}_j$ and $\boldsymbol{\xi}_j$, for j=1,2, respectively represent the decomposed components according to Eq. (1). $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ are basic reconstructed outputs of CAN (see Section 3.2). We call $\mathbf{x}_1$–to–$\mathbf{x}_2$ is an aging direction and vice versa, $\mathbf{x}_2$–to–$\mathbf{x}_1$ is a de-aging direction. Two single-hidden-layer neural networks as a bi-directional *bridge* are chosen to connect $\mathbf{A}_1$ and $\mathbf{A}_2$ to fit aging and de-aging process. We limit the age gap of each training image pair in T within a certain range according to different datasets. This is used to guarantee the aging and de-aging fitting process effective. To encode age-invariant features $\mathbf{I}_1$ and $\mathbf{I}_2$ from inputs $\mathbf{x}_1$ and $\mathbf{x}_2$, in our model, specifically we have two steps:

1. **Basic reconstruction:** This step respectively reconstructs the facial image inputs $\mathbf{x}_1$ and $\mathbf{x}_2$ independently by two auto-encoders to capture as much as main factors of inputs. Inputs are projected into a high-dimensional feature space in hidden layers.
2. **Transfer:** This step imposes constraints in the above feature space to nonlinearly decompose it into three feature subspaces: identity feature space which is age-invariant, age feature space which is identity-independent and a noise space.

The two steps above build our CAN model. We detail the two steps based on their cost functions in the following two sections.

3.2. Basic reconstruction

In this step, given a pair of facial images of the same person, we respectively reconstruct these two images by CAN. The cost function is defined as:

$$\min_{\theta_1} \quad \mathcal{L}_r = \frac{1}{2N}\sum_{i=1}^{N}(\| \mathbf{x}_1^i - \tilde{\mathbf{x}}_1^i \|_2^2 + \| \mathbf{x}_2^i - \tilde{\mathbf{x}}_2^i \|_2^2), \tag{2}$$

where parameters $\theta_1 = \{\mathbf{W}_j, \widehat{\mathbf{W}}_j, \mathbf{b}_j, \mathbf{c}_j\}$ where $\mathbf{W}_j = \{\mathbf{W}_{uj}, \mathbf{W}_{vj}, \mathbf{W}_{nj}\}$ and $\widehat{\mathbf{W}}_j = \{\widehat{\mathbf{W}}_{uj}, \widehat{\mathbf{W}}_{vj}, \widehat{\mathbf{W}}_{nj}\}$, for j=1,2, as shown in Fig. 2. $\tilde{\mathbf{x}}_1^i$ and $\tilde{\mathbf{x}}_2^i$ are outputs of auto-encoders to reconstruct the corresponding inputs $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$. Eq. (2) is a typical auto-encoder training objective, i.e., a squared error function. In the rest of this paper, we ignore the average processing (like $\frac{1}{2N}$ in Eq. (2)) to analyze the cost functions for simplicity.

Here we only choose the first term to analyze since the two terms in Eq. (2) are completely similar. A basic auto-encoder has two main

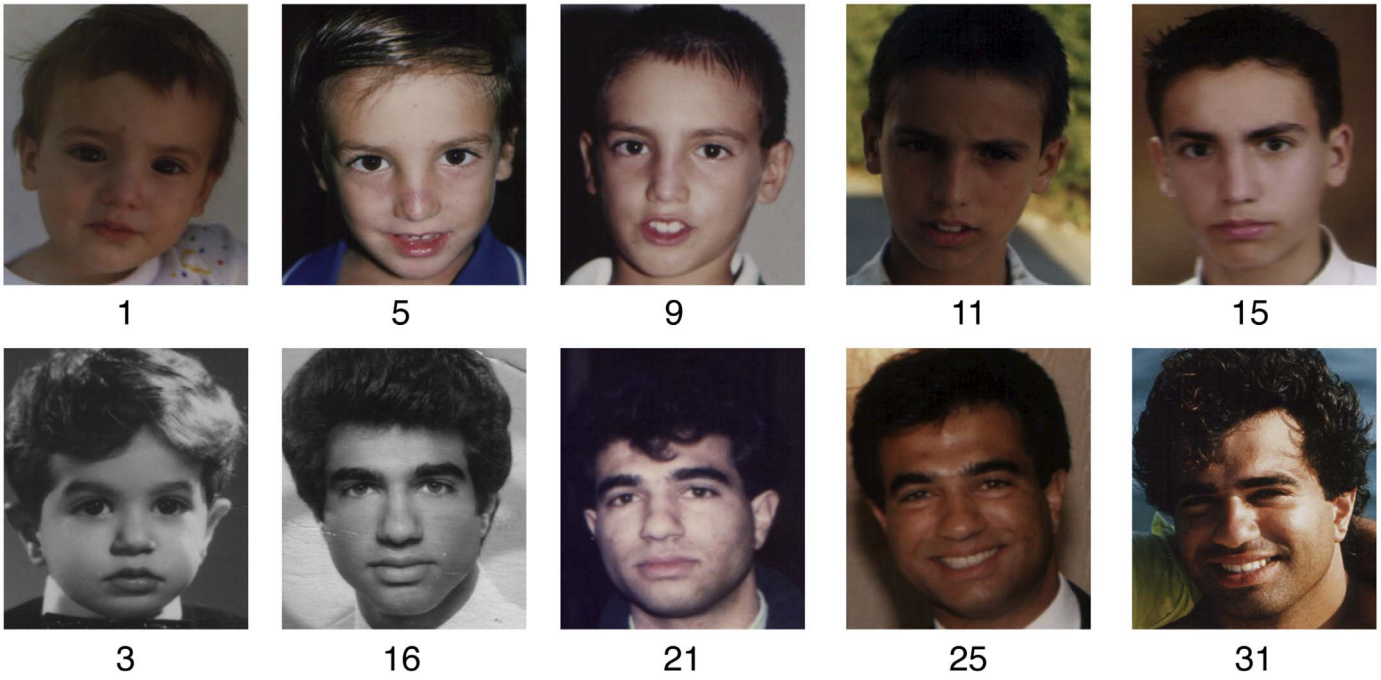

Fig. 1. Example images from FGNET. Images of the same row are of the same subject. The number at the bottom shows the age of the image.

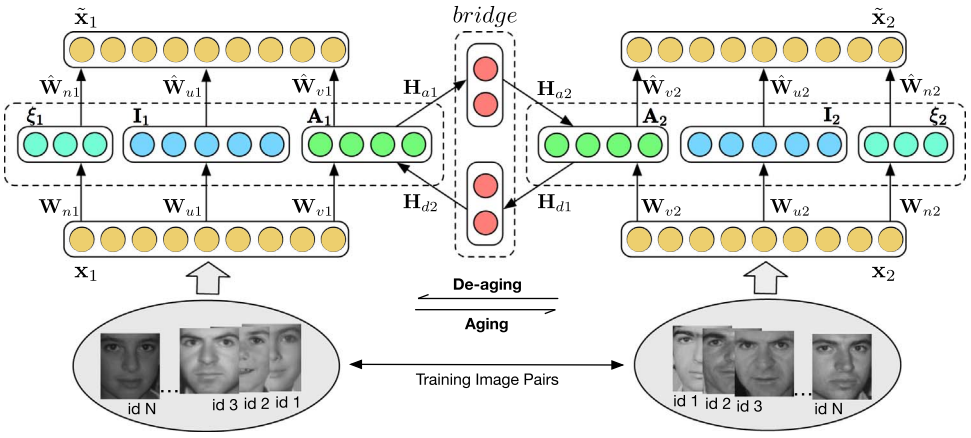

Fig. 2. The overview of CAN. CAN is composed of two identical auto-encoders and a *bridge* network. Given a pair of input images (**x1**, **x2**) of one person, first we leverage auto-encoders to reconstruct inputs to project them into a high-dimensional feature space in hidden layers. Second, we add constraints in the above feature space to decompose it into three components where ($\mathbf{I}_1$, $\mathbf{I}_2$) as identity features can be used as age-invariant representations for recognition and retrieval. Note here different id can refer to the same person. Details of CAN are described in Sections 3.2 and 3.3.

blocks called encoder and decoder. Input $\mathbf{x}_1^i$ can be encoded by a function $\mathbf{h}_1 = f_1(\mathbf{x}_1^i)$, which can be written as:

$$\mathbf{h}_1^i = f_1(\mathbf{x}_1^i) = s(\mathbf{W}_1\mathbf{x}_1^i + \mathbf{b}_1), \tag{3}$$

for $i = 1, 2, \ldots, N$. $\mathbf{W}_1 \in \mathbf{R}^{m\times n}$ is a weight matrix where m is the number of neurons in the hidden layer and $\mathbf{b}_1 \in \mathbf{R}^{m\times 1}$ is a hidden layer bias vector. $s(z) = (1 + e^{-z})^{-1}$ is a sigmoid function. $\mathbf{h}_1^i$ is the hidden layer representation.

In the decoding stage, $\mathbf{h}_1^i$ as the input is decoded by another function $\mathbf{g}_1$ to get $\tilde{\mathbf{x}}_1^i$:

$$\tilde{\mathbf{x}}_1^i = g_1(\mathbf{h}_1^i) = s_1(\widehat{\mathbf{W}}_1\mathbf{h}_1^i + \mathbf{c}_1), \tag{4}$$

where $\tilde{\mathbf{x}}_1^i$ is a reconstructed output to be close to the input $\mathbf{x}_1^i$. $\widehat{\mathbf{W}}_1 \in \mathbf{R}^{n\times m}$ is a weight matrix and $\mathbf{c}_1 \in \mathbf{R}^{n\times 1}$ is an output layer bias vector. $s_1(z) = z$ is an identity function (i.e., linear activation function). Minimizing this term can update $\{\mathbf{W}_1, \widehat{\mathbf{W}}_1, \mathbf{b}_1, \mathbf{c}_1\} \subset \theta_1$. Similarly, through minimizing the second term, we can update $\{\mathbf{W}_2, \widehat{\mathbf{W}}_2, \mathbf{b}_2, \mathbf{c}_2\} \subset \theta_1$. After solving Eq. (2) we fix the updated parameters θ_1 for step 2.

3.3. Transfer

After performing step 1, we impose constraints in the hidden layer representation $\mathbf{h}_j$ to decompose it into three feature subspaces: $\mathbf{I}_j$, $\mathbf{A}_j$ and $\boldsymbol{\xi}_j$, for j=1,2, as shown in Fig. 2. Below we formulate the cost function of this step as:

$$\min_{\theta_2}\mathcal{L}_t = \frac{1}{2N}\sum_{i=1}^{N}(\| \mathbf{A}_2^i - \widehat{\mathbf{A}}_2^i \|_2^2 + \| \mathbf{A}_1^i - \widehat{\mathbf{A}}_1^i \|_2^2 + \| \mathbf{I}_2^i - \mathbf{I}_1^i \|_2^2 + \| \mathbf{x}_2^i - \hat{\mathbf{x}}_2^i \|_2^2 + \| \mathbf{x}_1^i - \hat{\mathbf{x}}_1^i \|_2^2), \tag{5}$$

where parameters $\theta_2 = \{\mathbf{W}_{uj}, \widehat{\mathbf{W}}_{uj}, \mathbf{W}_{vj}, \widehat{\mathbf{W}}_{vj}, \mathbf{b}_{uj}, \mathbf{b}_{vj}, \mathbf{c}_j, \mathbf{H}_{aj}, \mathbf{H}_{dj}, \mathbf{b}_{aj}, \mathbf{b}_{dj}\}$, for j=1,2. $\widehat{\mathbf{A}}_2^i$ is an aging fitting output encouraged to be equal to the target older age feature $\mathbf{A}_2^i$. In the de-aging direction, $\widehat{\mathbf{A}}_1^i$ is encouraged to be equal to $\mathbf{A}_1^i$, the target younger age feature. $\mathbf{I}_1^i$ and $\mathbf{I}_2^i$ are identity features of the same person which are age-invariant. Here $\hat{\mathbf{x}}_2^i$ and $\hat{\mathbf{x}}_1^i$ we call them *transfer* reconstruction outputs to approximate inputs $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$, respectively.

Minimizing the first squared error term $\| \mathbf{A}_2^i - \widehat{\mathbf{A}}_2^i \|_2^2$ in Eq. (5) is

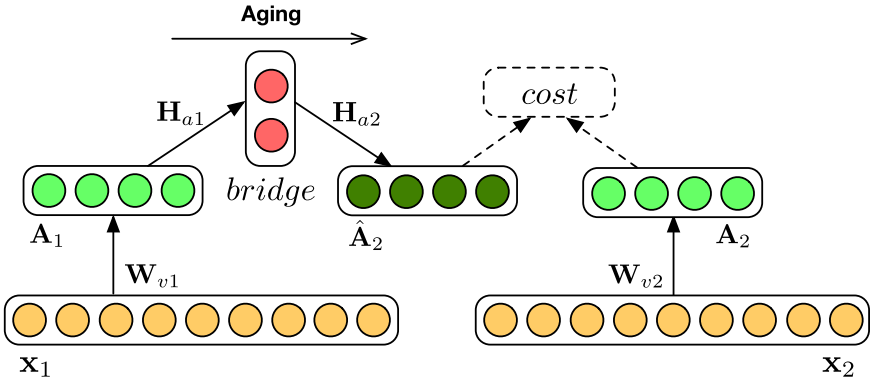

Fig. 3. Aging fitting neural network. $\hat{\mathbf{A}}_2$ is an aging fitting output encouraged to be equal to the target older age term $\mathbf{A}_2$. $\mathbf{A}_1$ is the younger age feature. We leverage *bridge* network to fit this aging process.

used to fit aging process between the two age features $\mathbf{A}_2^i$ and $\mathbf{A}_1^i$. We choose a single-hidden-layer neural network to connect $\mathbf{A}_2^i$ and $\mathbf{A}_1^i$ to fit this process because of the fact that any single-hidden-layer neural work can fit any complex smooth function [10]. Further, we observe that aging (and de-aging) process is a highly complex but smooth transform process. Optimizing this term in fact is to train an aging fitting neural network separated from CAN as shown in Fig. 3, and $\mathbf{A}_2^i$ can be expressed as follows:

$$\mathbf{A}_2^i = f_{v2}(\mathbf{x}_2^i) = s(\mathbf{W}_{v2}\mathbf{x}_2^i + \mathbf{b}_{v2}), \tag{6}$$

for $i = 1, 2, \ldots, N$, where f_{v2} is a nonlinear function forced to encode age feature from input $\mathbf{x}_2^i$. $\mathbf{W}_{v2} \in \mathbf{R}^{q\times n}$ is a weight matrix where q is age feature dimension (i.e., the number of neurons). $\mathbf{b}_{v2} \in \mathbf{R}^{q\times 1}$ is age feature bias vector.

Before continuing our analysis, we first define two functions called aging and de-aging function, $\mathcal{F}_a$ and $\mathcal{F}_d$, as:

$$\begin{cases} \mathcal{F}_a(\mathbf{z}) = f_{a2}(f_{a1}(\mathbf{z})) \\ \mathcal{F}_d(\mathbf{z}) = f_{d2}(f_{d1}(\mathbf{z})) \end{cases}, \tag{7}$$

where f_{aj} and f_{dj} are defined as:

$$\begin{cases} f_{aj}(\mathbf{z}) = s(\mathbf{H}_{aj}\mathbf{z} + \mathbf{b}_{aj}) \\ f_{dj}(\mathbf{z}) = s(\mathbf{H}_{dj}\mathbf{z} + \mathbf{b}_{dj}) \end{cases}, \tag{8}$$

for $j = 1, 2$, where $\mathbf{H}_{a1}, \mathbf{H}_{d1} \in \mathbf{R}^{k\times q}$ and $\mathbf{H}_{a2}, \mathbf{H}_{d2} \in \mathbf{R}^{q\times k}$ are weight matrices. $\mathbf{b}_{a1}, \mathbf{b}_{d1} \in \mathbf{R}^{k\times 1}$ are middle layer bias vectors. We make age feature bias vectors, $\mathbf{b}_{v1}$ and $\mathbf{b}_{v2}$, to be adaptive, i.e., used for both encoding age feature and fitting aging and de-aging process, so here $\mathbf{b}_{a2} = \mathbf{b}_{v2}$, $\mathbf{b}_{d2} = \mathbf{b}_{v1}$. k is the number of *bridge* neurons. Thus, as shown in Fig. 2, *bridge* networks can be formulated as $\mathcal{F}_a$ and $\mathcal{F}_d$ which are highly nonlinear due to the composite of two sigmoid functions. Note the input of $\mathcal{F}_a$ in our model is $\mathbf{A}_1^i$ for aging while $\mathcal{F}_d$ is of the older input $\mathbf{A}_2^i$ for de-aging. Now we can formulate $\hat{\mathbf{A}}_2^i$ according to Eq. (7) as:

$$\hat{\mathbf{A}}_2^i = \mathcal{F}_a(\mathbf{A}_1^i), \tag{9}$$

where $\mathbf{A}_1^i$ has a similar definition as $\mathbf{A}_2^i$ in Eq. (6). The second term has a similar analysis because it only fits from an opposite direction for de-aging. Therefore, to minimize the first two terms in Eq. (5) technically is encouraged to extract age-related information from inputs.

The third term is to make sure the error between the two encoded identity features of the same person $\mathbf{I}_1^i$ and $\mathbf{I}_2^i$ is small. This term is based on the observation that facial images of the same person contain stable identity feature that is age-invariant. Here $\mathbf{I}_j^i$ can be formulated as below:

$$\mathbf{I}_j^i = f_{ij}(\mathbf{x}_j^i) = s(\mathbf{W}_{ij}\mathbf{x}_j^i + \mathbf{b}_{ij}), \tag{10}$$

for $i = 1, 2, \ldots, N$, $j = 1, 2$. In the above equation, f_{ij} is an identity encoding function, $\mathbf{W}_{ij} \in \mathbf{R}^{p\times n}$ is a weight matrix where p is identity feature dimension and $\mathbf{b}_{ij} \in \mathbf{R}^{p\times 1}$ is identity feature bias vector. Minimizing $\| \mathbf{I}_2^i - \mathbf{I}_1^i \|_2^2$ is encouraged to extract common identity information from inputs of the same person. Further we will use $\mathbf{I}_1^i$ and $\mathbf{I}_2^i$ as age-invariant representations for face recognition and retrieval.

The fourth term is a *transfer* reconstruction squared error. Here we actually train a *transfer* reconstruction neural network separated from CAN as shown in Fig. 4. For the inputs $(\mathbf{x}_1^i, \mathbf{x}_2^i)$ of one person at different ages, our idea is to use the aging fitting output $\hat{\mathbf{A}}_2^i$ combined with $\mathbf{I}_2^i$ to reconstruct $\mathbf{x}_2^i$, the target older facial image input. We call this process as *transfer* reconstruction. Here $\hat{\mathbf{x}}_2^i$ is formulated as:

$$\hat{\mathbf{x}}_2^i = s_l(\widehat{\mathbf{W}}_{v2}\hat{\mathbf{A}}_2^i + \widehat{\mathbf{W}}_{u2}\mathbf{I}_2^i + \mathbf{c}_2), \tag{11}$$

for $i = 1, 2, \ldots, N$, where $\widehat{\mathbf{W}}_{v2} \in \mathbf{R}^{n\times q}$ and $\widehat{\mathbf{W}}_{u2} \in \mathbf{R}^{n\times p}$ are weight matrices. $s_l(z)$ is an identity function. Similarly, $\hat{\mathbf{x}}_1^i$ in the fifth term can be formulated as:

$$\hat{\mathbf{x}}_1^i = s_l(\widehat{\mathbf{W}}_{v1}\hat{\mathbf{A}}_1^i + \widehat{\mathbf{W}}_{u1}\mathbf{I}_1^i + \mathbf{c}_1). \tag{12}$$

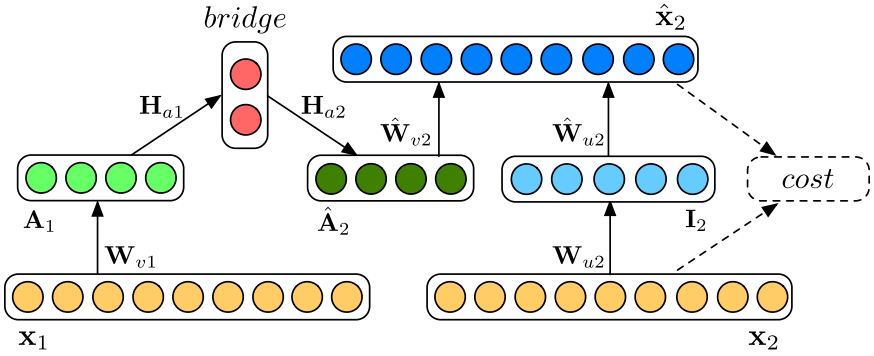

Fig. 4. Transfer reconstruction neural network. Given inputs $(\mathbf{x}_1, \mathbf{x}_2)$ of one person at different ages, we use aging fitting output $\hat{\mathbf{A}}_2$ combined with target identity feature $\mathbf{I}_2^i$ to reconstruct the older facial image input $\mathbf{x}_2$.

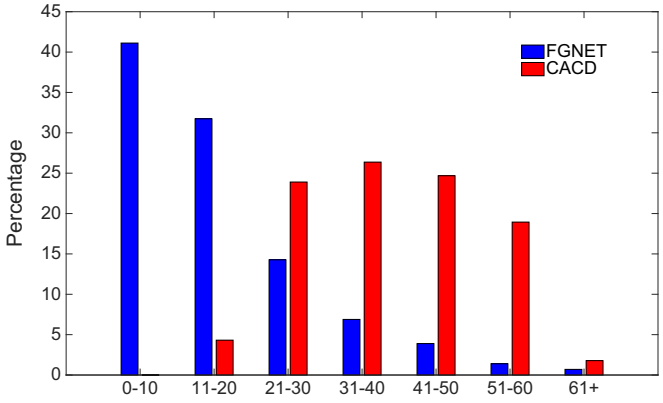

Fig. 5. Age range distribution (%) of FGNET and CACD.

Minimizing the last two terms in Eq. (5) can make as much as useful personal information concentrated on parameters θ_2. Combined with constraints brought by other terms in Eq. (5), we simultaneously separate identity-related and age-related information as we need. For the noises ξ_1 and ξ_2, we choose to separate them from inputs indirectly by our CAN training algorithm (see Section 3.4).

3.4. Training

Training CAN involves two steps as discussed above and we alternately perform these two training steps. We describe our training procedure in Algorithm 1. In *transfer* step, we add constraints on identity-related and age-related parameters $\mathbf{W}_{uj}, \mathbf{W}_{vj} \subset \mathbf{W}_j$, $\widehat{\mathbf{W}}_{uj}, \widehat{\mathbf{W}}_{vj} \subset \widehat{\mathbf{W}}_j$, $\mathbf{b}_{uj}, \mathbf{b}_{vj} \subset \mathbf{b}_j$, for j=1,2, to encode identity and age features. Combined with *basic reconstruction* step, this overall training, we separate other irrelevant information in noise-related parameters $\mathbf{W}_{nj} \subset \mathbf{W}_j$, $\widehat{\mathbf{W}}_{nj} \subset \widehat{\mathbf{W}}_j$, $\mathbf{b}_{nj} \subset \mathbf{b}_j$, for j=1,2. Therefore we indirectly encode noise ξ_1 and ξ_2 in hidden layers. Here to solve Eqs. (2) and (5), we adopt stochastic gradient descent (SGD) using standard back-propagation [45].

Algorithm 1. CAN training.

Input: training set $T = \{\mathbf{x}_1^i, \mathbf{x}_2^i\}(\mathbf{x}_1^i, \mathbf{x}_2^i \in \mathbf{R}^n, i = 1, 2, 3, \ldots, N)$; feature dimension p, q, r and the number of *bridge* neurons k; mini-batch size m' and iteration epoch *maxEpoch*; learning rate α.

Output: identity encoder parameters $\mathbf{W}_{uj}^*, \mathbf{b}_{uj}^*(j = 1, 2)$

Set t=1. Initialize $\mathbf{W}_j$, $\widehat{\mathbf{W}}_j$, $\mathbf{H}_{aj}$, $\mathbf{H}_{dj}(j = 1, 2) \sim \mathcal{N}(0, 10^{-4})$ and $\mathbf{b}_{aj}$, $\mathbf{b}_{dj}$, $\mathbf{b}_j$, $\mathbf{c}_j(j = 1, 2)$ to be all 0.
repeat
  Shuffle T.
  repeat
    Pick a mini-batch T' from T without overlapping.
    Compute $\mathcal{L}_r$ in Eq. (2).
    Update parameters θ_1 by solving Eq. (2).
    Compute $\mathcal{L}_t$ in Eq. (5).
    Update parameters θ_2 by solving Eq. (5).
  until T is looped over.
  $t = t + 1$.
until *maxEpoch* is met.

Since our CAN training is an unsupervised learning method, the extracted age-invariant features $\mathbf{I}_1$ and $\mathbf{I}_2$ in hidden layers are not discriminative, so they can not be directly used for face recognition and retrieval. Here as same as the strategy of [4,5], we employ PCA [46] on extracted $\mathbf{I}_1$ and $\mathbf{I}_2$ followed by LDA [11,47], a supervised dimension reduction technique, to make them more compressed and discriminative as the final age-invariant features for face recognition and retrieval.

3.5. Matching method

After CAN training and dimension reduction, we need the learned identity encoder parameters $\mathbf{W}_{uj}^*, \mathbf{b}_{uj}^*(j = 1, 2)$ and trained PCA and LDA transform matrices $\mathbf{M}_p, \mathbf{M}_l$ to obtain the final age-invariant features. Concretely, given a pair of probe and gallery facial image inputs $(\mathbf{x}_p, \mathbf{x}_g)$, according to Eq. (10), corresponding age-invariant features are computed as follows:

$$\mathbf{I}_p = \mathbf{M}_l^{\mathrm{T}}\mathbf{M}_p^{\mathrm{T}} f_{u1}^*(\mathbf{x}_p) = \mathbf{M}_l^{\mathrm{T}}\mathbf{M}_p^{\mathrm{T}} s(\mathbf{W}_{u1}^*\mathbf{x}_p + \mathbf{b}_{u1}^*), \tag{13}$$

$$\mathbf{I}_g = \mathbf{M}_l^{\mathrm{T}}\mathbf{M}_p^{\mathrm{T}} f_{u2}^*(\mathbf{x}_g) = \mathbf{M}_l^{\mathrm{T}}\mathbf{M}_p^{\mathrm{T}} s(\mathbf{W}_{u2}^*\mathbf{x}_g + \mathbf{b}_{u2}^*), \tag{14}$$

where the superscript $^{\mathrm{T}}$ means a transposition of a matrix. Then we use cosine distance to compute matching scores between $\mathbf{I}_p$ and $\mathbf{I}_g$ for age-invariant face recognition and retrieval.

4. Experiment

4.1. Datasets

We evaluate our approach on three public aging face datasets: FGNET [12], CACD [13] and CACD-VS [14]. FGNET contains 1,002 images of 82 different people, with each one has about 13 images on average taken at different ages from 0 to 69. CACD is a new large-scale dataset collected from the Internet which consists 163,446 face images of 2,000 people with age ranged from 16 to 62. To the best of our knowledge, CACD is the largest public available face aging dataset. Compared to CACD, FGNET has larger age gap and more younger images, but CACD has a larger number of images and more images at other ages. Fig. 5 shows the age range distribution of these two datasets.

Further, we conduct an experiment on CACD-VS, a subset of CACD, for face verification. CACD-VS dataset contains 2,000 positive pairs and 2,000 negative pairs and is carefully annotated by checking both of the associated image and surrounding web contents.

In our problem, all facial images are preprocessed as follows: (1) convert the images into gray ones if they are RGB images; (2) detect the locations of the faces in the images using Viola-Jones face detector [48] and locate the 83 landmarks using Face++ API [49]; (3) align the images to make their eyes located at the same horizontal positions; (4) crop the images to remove the background and hair region; (5) rescale them by bicubic interpolation and reshape them into one-dimension vector. All the data are then mapped into $[0, 1]$ and normalized to have zero mean.

4.2. Parameters setting

In our approach, there are several hyper-parameters to select: input dimension n, identity feature dimension p, age feature dimension q, noise-related feature dimension r, the number of *bridge* neurons k in CAN and dimension of PCA [46] and LDA [11,47]. These hyper-parameters setting is given in Table 1. For FGNET and CACD datasets, input dimension n is 35×32 and in CAN training, we choose a fixed learning rate α=0.0001, mini-batch size m' to be 10 to perform SGD. For CACD-VS, we use the same parameters setting in CACD. For the

Table 1
The parameters setting in our experiments.

Dataset		FGNET	CACD
CAN	p	2100	2800
	q	600	800
	r	300	400
	k	500	800
Dimension reduction	PCA	400	500
	LDA	100	120

three datasets, we respectively set iteration epoch *maxEpoch* to 1,000, 500 and 800. All parameters updating use momentum of 0.9.

4.3. Experiment on FGNET dataset

FGNET is a challenging face aging dataset because it is relatively small and suffers from other significant variations such as pose, illumination and expression. Some examples of them are shown in Fig. 1. Following the training and testing split scheme in [6], we use leave-one-image-out strategy for performance evaluation. In each cross, we use all the remaining face images choosing pairs from them to form our training set (1,800 training image pairs for each cross and different pairs can refer to the same person). We constrain age gap of each image pair to be less than 10 years. All the training data are used to learn PCA and LDA subspaces in each cross.

4.3.1. Evaluation metrics

In our experiment on FGNET, we use leave-one-image-out strategy with rank-k identification rates for performance evaluation. Specifically, we leave one image as the test sample and train the model by using the remaining 1,001 images from which training pairs are selected. We repeat this procedure 1,002 times and took the average as the final identification rates. Cosine similarity is used to compute matching scores between the test example and remaining images. For rank-k, we sort the matching results from top-1 to top-k for each test example. Then we can get rank-k identification rates after averaging these results.

4.3.2. Parameters exploration

There are some parameters influencing the performance of our approach: the number of hidden layer neurons m of each auto-encoder in CAN where $m = p + q + r$, PCA dimension d_p and LDA dimension d_l. We use rank-k identification rates on FGNET to decide these parameters.

For the number of hidden layer neurons m, we run experiments from 1,000 to 4,000. For each choice of m, the parameters p, q and r are selected with exhaustive research. We give them in Table 2 and keep the number of *bridge* neurons k to be 500. Here we use raw age-invariant feature $\mathbf{I}_1$ and $\mathbf{I}_2$ (extracted from hidden layers) directly for identification to choose m. Theoretically, more hidden layer neurons means more complex encoding functions we can learn and can catch more useful information from inputs. Fig. 6(a) shows face recognition performance on FGNET of different choices of m. As we expect, the less neurons, the worse performance. We can observe that when m=3,000, the performance is slightly better than that of m=4,000, hence we choose m=3,000.

Different choices of PCA and LDA parameters are further investigated based on the above setting with m=3,000 in Table 2. For PCA dimension d_p, we select it from 100 to 1,000 and for LDA dimension d_l we select it from 60 to 300. And we run experiments to seek the best combination of them using rank-1 identification rates on FGNET as metric. Results of different PCA dimensions and LDA dimensions of rank-1 recognition rates are given in Table 3. Here we choose $(d_p, d_l) = (400, 100)$ as the best settings for dimension reduction and testing.

Table 2
The setting of age feature dimension p, identity feature dimension q, noise-related feature dimension r based on different choices of the number of hidden layer neurons m.

m	1000	2000	3000	4000
p	700	1400	2100	2800
q	200	400	600	800
r	100	200	300	400

4.3.3. Effects of dimension reduction strategies

We also study the performance of our approach with different dimension reduction strategies. We test our model on FGNET with PCA only, LDA only and PCA+LDA applied on raw age-invariant features $\mathbf{I}_1$ and $\mathbf{I}_2$. Concretely, for PCA only, we follow the above setting $d_p = 400$, for LDA only we tune d_l to 200, and for PCA+LDA, we set $(d_p, d_l) = (400, 100)$ as above. Cumulative Match Characteristic (CMC) curves on FGNET of different dimension reduction strategies are shown in Fig. 6(b). We have several observations. First, raw feature performs badly mainly due to lack of supervised information. Second, there are significant performance improvements after LDA applied, whether or not PCA is applied, and rank-1 identification rate based on raw features can be improved from 38.46–75.25% after only LDA applied. This demonstrates the effectiveness of supervised learning method combined with our unsupervised CAN model. Finally, PCA technique can further improve our performance. Therefore, we use raw feature with PCA+LDA strategy in our following experiments.

4.3.4. Comparison with state-of-the-art algorithms

We compare our approach with state-of-the-art algorithms including: (1) a generative model to build face aging space for age-invariant face recognition [1]; (2) a discriminative model [6]; (3) hidden factor analysis [4], a linear factor analysis method for face recognition; (4) a discriminative method using a maximum entropy feature descriptor based on identity factor analysis [5]; (5) a deep face recognition framework called latent factor guided convolutional neural network (LF-CNN) [32]. (6) a CNN baseline model [32] with same networks as LF-CNNs without latent identity analysis (LIA). Comparative results are shown in Table 4.

Among all the compared algorithms in Table 4, it can be seen that our approach obtains competitive results. As we can see, our nonlinear factor analysis method with CAN is superior over other linear factor analysis methods in [4,5]. The performance of our approach is inferior to that of LF-CNNs [32] which has the top performance. One possible reason is that although we leverage LDA technique to make the identity features more discriminative, the proposed unsupervised CAN model is still less discriminative than the supervised LF-CNNs. Another possible reason is that our CAN adopts a shallow neural network architecture. Generally speaking, the performance of shallow networks is worse than that of deep networks.

Further, it is desirable to investigate our approach on different age groups in FGNET. Following the age groups setting in [32], rank-1 recognition rates of our approach on different age groups in FGNET are given in Table 5. From Table 5, we can see that our approach still yields good performance. This proves the effectiveness of our approach.

Finally, some failed retrieval results in FGNET are given in Fig. 7. We can see some incorrect rank-1 results are even more similar to the probe images than the corresponding ground-truth images. On the other hand, face recognition fails due to some other variations like illumination, expression, etc.

4.3.5. Effects of aging and de-aging operator

We leverage transfer reconstruction neural network (see Fig. 4) to investigate aging and de-aging operator of CAN. For aging operator, given an image pair as input $(\mathbf{x}_1, \mathbf{x}_2)$, we replace the decomposed age feature $\mathbf{A}_2$ of $\mathbf{x}_2$ with aging fitting output $\widehat{\mathbf{A}}_2$ to reconstruct $\mathbf{x}_2$. The goodness of the age feature from the hidden layer can be shown from the output of the reconstructed result $\hat{\mathbf{x}}_2$. For de-aging operator, the process is completely similar from an opposite direction. Here we use CAN trained with CACD dataset (see Section 4.4) to visualize some reconstructed results in FGNET. They are shown in Fig. 8. From the results, we can see the reconstructions are similar to the ground-truth images. This intuitively demonstrates the effectiveness of CAN to fit complex nonlinear aging and de-aging process.

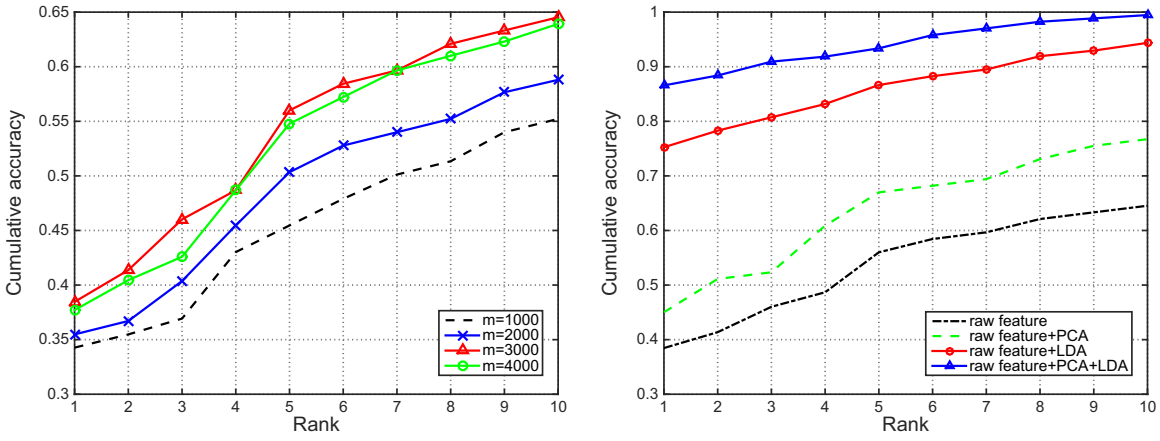

Fig. 6. (a) Cumulative Match Characteristic (CMC) curves on FGNET with different choices of the number of hidden layer neurons m of CAN. (b) CMC curves on FGNET with different dimension reduction strategies.

Table 3
Rank-1 recognition rates (%) of different PCA dimensions and LDA dimensions on FGNET.

d_p	d_l						
	60	100	140	180	220	260	300
100	79.0	–	–	–	–	–	–
200	83.5	82.5	80.8	78.1	–	–	–
300	83.9	86.0	82.9	80.3	78.3	71.4	–
400	82.2	**86.5**	85.4	80.6	77.0	71.7	65.4
500	81.4	84.8	82.0	78.9	76.5	72.2	66.4
600	80.6	80.7	81.4	78.4	75.4	74.9	69.5
700	78.3	80.3	80.6	77.5	72.0	73.6	71.3
800	71.7	76.3	78.5	74.0	71.4	70.3	72.9
900	68.2	70.1	73.3	72.4	70.9	67.4	69.3
1000	63.1	66.2	69.7	70.8	67.2	65.9	65.7

Table 4
Rank-1 recognition rates of our approach compared with state-of-the-art algorithms on FGNET.

Algorithms	Recognition rate (%)
Park et al [1]	37.4
Li et al. [6]	47.5
HFA [4]	69.0
MEFA [5]	76.2
CNN-baseline [32]	84.4
LF-CNNs [32]	88.1
Our approach	**86.5**

Table 5
Performance of our approach compared with state-of-the-art algorithms on different age groups in FGNET.

Age group	Amount	CNN-baseline (%)	LF-CNNs (%)	Ours (%)
0–4	193	51.81	60.10	60.27
5–10	218	84.86	88.53	87.39
11–16	201	91.04	94.03	92.63
17–24	182	94.51	97.80	95.47
25–69	208	99.04	99.52	98.01

4.4. Experiment on CACD dataset

In this experiment, we conduct a face retrieval experiment on CACD [13], the newly largest public available face aging dataset. CACD dataset includes varying illumination, pose variation and makeup.

We follow the experimental settings in [13]. In CACD, 120 celebrities with rank 3–5 are chosen as test sets where images taken at 2013 are used as query images. The remaining images are split into three subsets respectively taken in 2004–2006, 2007–2009 and 2010–2012 as database images. In training, for each one of the remaining 1,880 celebrities in CACD, there are about 80 images taken at different years while the age gap is about 0–10 years. For these remaining images, we select 20 image pairs of each person to aggregate them as training set (37,600 pairs). Note that the use of makeup in CACD may confound the age of an individual. In order to avoid the impact brought by this on our algorithm, we carefully check the corresponding image contents and age labels to get our training set. All training images are then used to learn PCA+LDA subspaces. Age gap of each training image pair is constrained between 2 and 7 years.

4.4.1. Evaluation metrics

In our experiment on CACD, we use mean average precision (MAP) as evaluation metrics. Cosine distance is used to compute the similarity of two images. Concretely, let $q_i \in Q$ be the query images and Q is the query database. For q_i, the positive images can be expressed as $Y_1, Y_2, \ldots, Y_{m_i}$. We define E_{ic} as the retrieval results of q_i in a descending order, from the top to Y_c. We first give average precision (AP) of q_i as below:

$$AP(q_i) = \frac{1}{m_i} \sum_{c=1}^{m_i} Precision(E_{ic}), \tag{15}$$

where $Precision(E_{ic})$ means the ratio of positive images in E_{ic}. Then the MAP of Q can be computed as:

$$MAP(Q) = \frac{1}{|Q|} \sum_{i=1}^{|Q|} AP(q_i), \tag{16}$$

which is the average of the AP of all query images.

4.4.2. Comparison with state-of-the-art algorithms

We compare our approach with state-of-the-art algorithms including HFA [4], CARC [14] and a generalized similarity model [36] (GSM-1 and GSM-2). Compared with GSM-1, GSM-2 only uses more training samples. Fig. 9 reports the comparative results. All methods in Fig. 9 are tuned to the best settings according to their papers. The results in Fig. 9 show that our approach outperforms the others in all three subsets. Note that compared with HFA [4], CARC [14] and GSM-1 [36] on the subset with small age gap, both our method and GSM-2 [36] can achieve competitive performance on the subset with large age gap. This confirms the superiority of our approach.

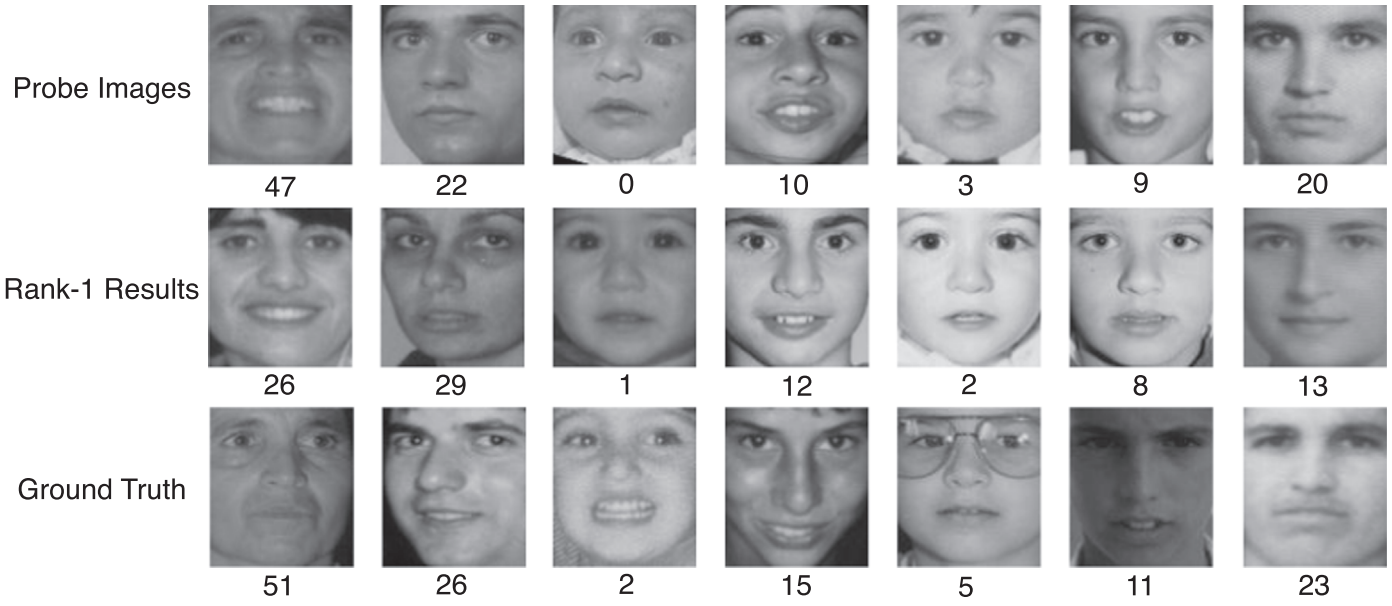

Fig. 7. Some failed retrieval results in FGNET. The first row shows the probe faces and the second row shows the incorrect rank-1 retrieval results using our approach. The third row presents the ground-truth images corresponding to the probes.

4.5. Experiment on CACD-VS dataset

CACD-VS dataset contains 4,000 images pairs from 2,000 celebrities. Following the configuration in [14] for face verification, we split CACD-VS into ten folds and each fold has 400 images pairs (200 positive pairs and 200 negative pairs) from 200 celebrities. We use one fold for testing and the other nine folds for training. We repeat our experiment on each of the ten folds and report average results. Concretely, for each run, we use the other nine folds (3,600 image pairs) to train CAN and learn PCA+LDA subspaces. After we get the identity feature for each image, cosine similarity is used to compute matching scores between pairs. The optimal classification threshold is decided by the nine training folds. Performance of our method compared with state-of-the-art algorithms is reported in Table 6.

From the results reported in Table 6, although our method significantly improves verification accuracy from 85.7–92.3% compared with human average performance, combining the decisions from multiple human can get a higher accuracy of 94.2%. It proves that there is still a gap for our method to achieve human performance. We also add two general deep face recognition methods for comparison, Deepface [50] and DeepID2 [51]. The result of Deepface is borrowed from [36]. DeepID2 model is pretrained with CACD dataset. As seen in

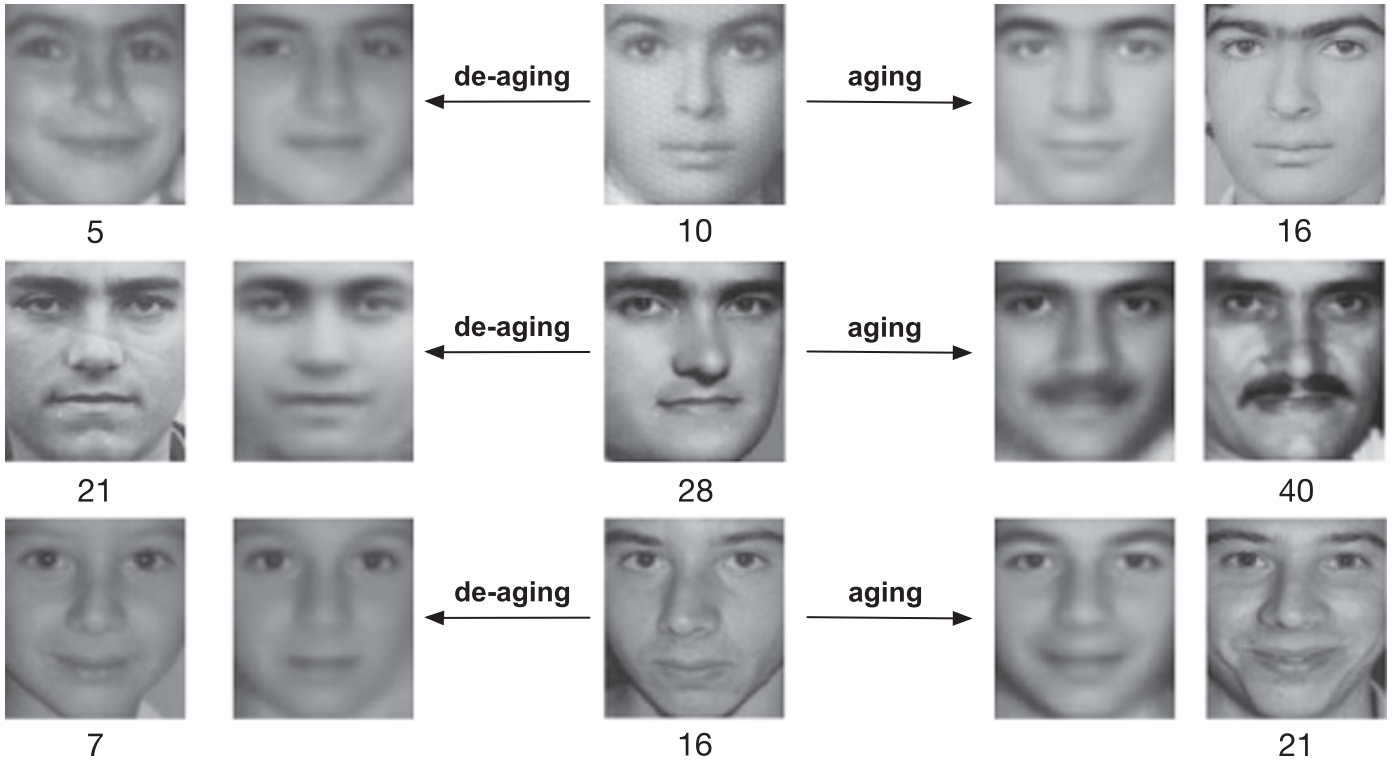

Fig. 8. Some aging and de-aging visualization results in FGNET. Each row represents the same person. The second and fourth column show the reconstructed outputs. The first and last column show the ground-truth images.

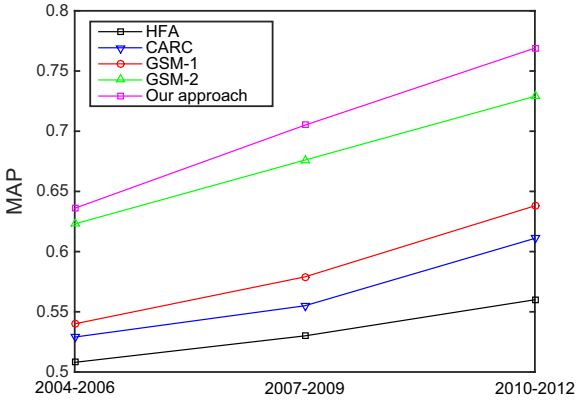

Fig. 9. Face retrieval performance in terms of MAP of our approach compared with state-of-the-art algorithms on CACD.

Table 6
Verification accuracy on the CACD-VS dataset.

Method	Accuracy (%)
HD-LBP [52]	81.6
HFA [4]	84.4
CARC [14]	87.6
Deepface [50]	85.4
DeepID2 [51]	87.2
DCNN+LBPH [39]	89.5
Human, Average	85.7
Human, Voting (2015)	94.2
LF-CNNs [32]	98.5
GSM [36]	89.8
Our approach	**92.3**

Table 6, our method still outperforms them. This further demonstrates the specific effectiveness of CAN on face verification with aging variations.

5. Conclusions

In this paper, we propose coupled auto-encoder networks (CAN) and a nonlinear factor analysis method, to address age-invariant face recognition and retrieval problem. Through CAN, we can nonlinearly separate identity feature to be age-invariant from one given face image. Experiments on FGNET, CACD and CACD-VS confirm the effectiveness of our approach.

In the future, we will attempt to incorporate supervised information in CAN and refine our networks architecture. Cross-database evaluation will be investigated. We will also extend our CAN model to tackle face recognition problems with other variations like expression, illumination and pose.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61375038) and Applied Basic Research Programs of Sichuan Science and Technology Department (2016JY0088).

References

[1] U. Park, Y. Tong, A.K. Jain, Age-invariant face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 947–954.
[2] J.-X. Du, C.-M. Zhai, Y.-Q. Ye, Face aging simulation and recognition based on nmf algorithm with sparseness constraints, Neurocomputing 116 (2013) 250–259.
[3] Z. Li, D. Gong, X. Li, D. Tao, Aging face recognition: a hierarchical learning model based on local patterns selection, IEEE Trans. Image Process. 25 (5) (2016) 2146–2154.
[4] D. Gong, Z. Li, D. Lin, J. Liu, X. Tang, Hidden factor analysis for age invariant face recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2872–2879.
[5] D. Gong, Z. Li, D. Tao, J. Liu, X. Li, A maximum entropy feature descriptor for age invariant face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5289–5297.
[6] Z. Li, U. Park, A.K. Jain, A discriminative model for age invariant face recognition, IEEE Trans. Inf. Forensics Secur. 6 (3) (2011) 1028–1037.
[7] H. Ling, S. Soatto, N. Ramanathan, D.W. Jacobs, Face verification across age progression using discriminative methods, IEEE Trans. Inf. Forensics Secur. 5 (1) (2010) 82–91.
[8] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[9] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[10] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inf. Theory 39 (3) (1993) 930–945.
[11] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[12] FG-NET Aging Database, ⟨http://www.fgnet.rsunit.com⟩.
[13] B.-C. Chen, C.-S. Chen, W.H. Hsu, Cross-age reference coding for age-invariant face recognition and retrieval, in: Computer Vision–ECCV 2014, Springer, 2014, pp. 768–783.
[14] B.-C. Chen, C.-S. Chen, W.H. Hsu, Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset, IEEE Trans. Multimed. 17 (6) (2015) 804–815.
[15] A. Montillo, H. Ling, Age regression from faces using random forests, in: Proceedings of the Image Processing (ICIP), 2009 16th IEEE International Conference on, IEEE, 2009, pp. 2465–2468.
[16] G. Guo, G. Mu, Y. Fu, T.S. Huang, Human age estimation using bio-inspired features, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 112–119.
[17] Y. Fu, T.S. Huang, Human age estimation with regression on discriminative aging manifold, IEEE Trans. Multimed. 10 (4) (2008) 578–584.
[18] S.K. Zhou, B. Georgescu, X.S. Zhou, D. Comaniciu, Image based regression using boosting method, in: Proceedings of the Computer Vision, 2005, ICCV 2005, Tenth IEEE International Conference on, vol. 1, IEEE, 2005, pp. 541–548.
[19] S. Yan, H. Wang, X. Tang, T.S. Huang, Learning auto-structured regressor from uncertain nonnegative labels, in: Proceedings of the Computer Vision, 2007, ICCV 2007, IEEE 11th International Conference on, IEEE, 2007, pp. 1–8.
[20] J. Wang, Y. Shang, G. Su, X. Lin, Age simulation for face recognition, in: Proceedings of the Pattern Recognition, 2006, ICPR 2006, 18th International Conference on, vol. 3, IEEE, 2006, pp. 913–916.
[21] N. Ramanathan, R. Chellappa, Face verification across age progression, IEEE Trans. Image Process. 15 (11) (2006) 3349–3361.
[22] G. Guo, Y. Fu, C.R. Dyer, T.S. Huang, Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 17 (7) (2008) 1178–1188.
[23] X. Geng, Z.-H. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2234–2240.
[24] Y.H. Kwon, N.D.V. Lobo, Age classification from facial images, in: Computer Vision and Pattern Recognition, 1994, Proceedings CVPR'94, 1994 IEEE Computer Society Conference on, IEEE, 1994, pp. 762–767.
[25] A. Lanitis, C. Draganova, C. Christodoulou, Comparing different classifiers for automatic age estimation, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 34 (1) (2004) 621–628.
[26] J. Lu, V.E. Liong, J. Zhou, Cost-sensitive local binary feature learning for facial age estimation, IEEE Trans. Image Process. 24 (12) (2015) 5356–5368.
[27] J. Suo, X. Chen, S. Shan, W. Gao, Learning long term face aging patterns from partially dense aging databases, in: Proceedings of the Computer Vision, 2009 IEEE 12th International Conference on, IEEE, 2009, pp. 622–629.
[28] A. Lanitis, C.J. Taylor, T.F. Cootes, Toward automatic simulation of aging effects on face images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 442–455.
[29] J. Suo, S.-C. Zhu, S. Shan, X. Chen, A compositional and dynamic model for face aging, IEEE Trans. Pattern Anal. Mach. Intell. 32 (3) (2010) 385–401.
[30] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, Y. Miyake, Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin, ACM Trans. Graph. (TOG) 22 (3) (2003) 770–779.
[31] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2041–2056.
[32] Y. Wen, Z. Li, Y. Qiao, Age invariant deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[33] Y. Li, G. Wang, L. Lin, H. Chang, A deep joint learning approach for age invariant face verification, in: Computer Vision, Springer, 2015, pp. 296–305.
[34] S. Bianco, Large Age-gap Face Verification by Feature Injection in Deep Networks, arXiv preprint arXiv:1602.06149.
[35] H. El Khiyari, H. Wechsler, et al., Face recognition across time lapse using convolutional neural networks, J. Inf. Secur. 7 (03) (2016) 141.
[36] L. Lin, G. Wang, W. Zuo, F. Xiangchu, L. Zhang, Cross-domain Visual Matching Via Generalized Similarity Measure and Feature Learning.
[37] L. Liu, C. Xiong, H. Zhang, Z. Niu, M. Wang, S. Yan, Deep aging face verification with large gaps, IEEE Trans. Multimed. 18 (1) (2016) 64–75.
[38] J. Lu, V.E. Liong, G. Wang, P. Moulin, Joint feature learning for face recognition,

IEEE Trans. Inf. Forensics Secur. 10 (7) (2015) 1371–1383.
[39] H. Zhai, C. Liu, H. Dong, Y. Ji, Y. Guo, S. Gong, Face verification across aging based on deep convolutional networks and local binary patterns, in: Intelligence Science and Big Data Engineering, Image and Video Data Engineering, Springer, 2015, pp. 341–350.
[40] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[41] Y. Bengio, Learning deep architectures for ai, Found. Mach. Learn. 2 (1) (2009) 1–127.
[42] M. Kan, S. Shan, H. Chang, X. Chen, Stacked progressive auto-encoders (spae) for face recognition across poses, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1883–1890.
[43] Y. Liu, X. Hou, J. Chen, C. Yang, G. Su, W. Dou, Facial expression recognition and generation using sparse autoencoder, in: Smart Computing (SMARTCOMP), 2014 International Conference on, IEEE, 2014, pp. 125–130.
[44] Y. Zhang, R. Liu, S. Zhang, M. Zhu, Occlusion-robust face recognition using iterative stacked denoising autoencoder, in: Neural Information Processing, Springer, 2013, pp. 352–359.
[45] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[46] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Computer Vision and Pattern Recognition, 1991, Proceedings CVPR'91, IEEE Computer Society Conference on, IEEE, 1991, pp. 586–591.
[47] X. Wang, X. Tang, A unified framework for subspace face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1222–1228.
[48] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001, CVPR 2001, Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, IEEE, 2001, pp. I–511.
[49] Megvii: Face++, ⟨http://www.faceplusplus.com⟩. (Accessed 7-3-2014).
[50] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[51] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Advances in Neural Information Processing Systems, 2014, pp. 1988–1996.
[52] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3025–3032.

Chenfei Xu received the B.S. degree in 2014, from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, where he is currently pursuing the M.S. degree. His research interests include deep learning and computer vision.

Qihe Liu received the Ph.D. degree in computer science from the University of Electronic Science and Technology of China in 2005. He is currently an Associate Professor of Electronic Science and Technology of China. His current research interests include machine olfaction, data mining and computer vision. He has authored or co-authored over 20 research papers published in international journals and conference proceedings.

Mao Ye received the Ph.D. degree in mathematics from Chinese University of Hong Kong, in 2002. He is currently a professor and Director of CVLab at University of Electronic Science and Technology of China. His current research interests include machine learning and computer vision. In these areas, he has published over 70 papers in leading international journals or conference proceedings. He is an associated editor of the journal "Engineering Applications of Artificial Intelligence".