Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition

Tianyue Zheng, Weihong Deng, Jiani Hu

Beijing University of Posts and Telecommunications

No.10 Xitucheng Road, Haidian District, Beijing, China

2231135739@qq.com, whdeng@bupt.edu.cn, 40902063@qq.com

Abstract

While very promising results have been shown on face recognition related problems, age-invariant face recognition still remains a challenge. Facial appearance of a person changes over time, which results in significant intra-class variations. In order to address this problem, we propose a novel deep face recognition network called age estimation guided convolutional neural network (AE-CNN) to separate the variations caused by aging from the person-specific features which are stable. The carefully designed CNN model can learn age-invariant features for face recognition. To the best of our knowledge, this is the first attempt to use age estimation task for obtaining age-invariant features. Extensive results on two well-known public domain face aging datasets: MORPH Album 2 and CACD show the effectiveness of the proposed approach.

1. Introduction

Age-invariant face recognition has received increasing attention due to its wide range of application scenarios such as finding missing children after years, identifying criminals using photos taken many years ago and verifying passport.

In spite of the great advancement in face related works in recent years, age-invariant face recognition is still a challenging problem in real world applications. The major difficulty of the problem is that face appearance of a person changes greatly during the aging process. Those changes are various in different age periods and cause significant intra-class variations, as shown in Figure 1.

A traditional approach to tackle age-invariant face recognition problem is to synthesis face to match target age and then perform recognition [5, 22]. They try to construct a 2D/3D model to compensate for the shape changes which cause large intra-class variations and degrade the performance of face recognition. However, these generative models need strong parametric assumptions and accurate age

Figure 1. The cross-age images for one subject which show the large intra-class variations due to aging process.

labels. Due to the complexity in modeling aging process, these approaches are also computationally expensive, so they are not stable in real-world cross-age face recognition.

Recently, discriminative methods are proposed to handle age-invariant face problem [8, 9, 15, 16, 18]. For example, Ling *et al*. [18] use gradient orientation pyramid (GOP) as feature and use the support vector machine (SVM) as classifier for face recognition. Many discriminative approaches aim to design an appropriate feature and an effective matching system. However, the features they design still contain age information and are not specific for age-invariant face recognition problem. To separate the person-specific identity factor from age factor, Gong *et al*. [8, 9] propose the hidden factor analysis method (HFA). The facial image of a person can be expressed as a combination of an identity-specific component that is stable over the aging process and other component that reflects the aging effect, then the identity-specific component is used for age-invariant face recognition.

Motivated by the ability of convolutional neural network (CNN) to learn latent representations from the input and the fact that CNN has been successfully applied to many face-related problems, we use CNN to obtain features. Inspired by the belief that the face image is a combination of an age-specific component and an identity-specific component in [8], we expect that the resulting deep feature reduces

the variations caused by aging process as much as possible. Ideally, we want the feature containing only identity-related component to perform well in cross-age face recognition. However, the whole deep feature trained by large scale face images using softmax loss function is unavoidable to contain both the latent identity factor such as gender and the age factor such as wrinkles, so the whole feature is not a good choice considering the characteristic of age-invariant face recognition problem. We first utilize age estimation task to get age feature and then obtain age-specific factor from age feature, thus we can get latent identity-related feature by removing the age-specific factor contained in the whole feature.

In this paper, we propose an age estimation task guided face recognition framework to learn age-invariant features. Given training data with age label and identity label, AE-CNN achieves age estimation task and face recognition task at the same time. We choose a well-designed CNN to handle the problem, the first fully connected layer outputs the whole feature which includes age-related factor. Then, age feature is obtained by adding a fully connected layer after the whole feature layer using softmax loss function for age estimation. Another fully-connected layer is added after age feature to get age-specific factor. The age-invariant feature is obtained by subtracting age factor from the whole feature and is used to achieve cross-age face recognition task with softmax loss function.

The major contributions of this paper are summarized as follows:

- A new model for age-invariant face recognition problem is proposed based on convolutional neural network.

- We propose a new method to obtain age-invariant feature by subtracting age factor got by age estimation which is identity-independent from the whole feature. To the best of our knowledge, it is the first work to show the effectiveness of age estimation in obtaining age-invariant feature and achieve comparative results.

- Extensive experiments have shown that the proposed AE-CNN outperforms state-of-the-art on the two public facing aging datasets (MORPH Album 2 [23] and CACD [2]).

The rest of this paper is organized as follows. Section 2 discusses related works. Section 3 introduces the formulation and details the proposed AE-CNN. Section 4 presents the experimental results. Finally, Section 5 concludes this paper.

2. Related Works

Most age-related works focus on age simulation and age estimation which includes exact age estimation and age group estimation. Works that explicitly aim to solve age-invariant face recognition problem are limited. Traditional approaches roughly fall into two categories: generative approaches and discriminative approaches. Generative methods in [5, 7, 12] try to synthesis the input image to match the target image by constructing a 2D or 3D model to compensate for the aging process and then achieve face recognition. These approaches need accurate age labels and strong parametric assumptions which lead to unrealistic synthesis results. So they do not achieve good results in real-world age-invariant face recognition.

Some discriminative approaches are proposed and draw increasing attentions [11, 14–16, 18, 21]. Li et al. [16] use scale invariant feature transform (SIFT) [6] and multi-scale local binary pattern (LBP) [20] as feature and a variation of random subspace LDA approach (RS-LDA) [24] to do face recognition. Gong et al. [8] propose a method called hidden factor analysis (HFA). The appearance can be modeled as a combination of an identity factor that is age-invariant and an age factor affected by the aging process, and the method tries to separate the age-invariant component. They adopt Expectation Maximization (EM) algorithm to estimate model parameters and further they propose a maximum entropy feature [9] with separating the person-specific feature to improve the method. Li et al. [15] propose local pattern selection (LPS) as a new feature descriptor for cross-age face recognition.

Data-driven approach has also been used. Chen et al. [2] propose a coding framework called Cross-Age Reference Coding (CARC). The method is based on the assumption that if people look alike when they are young, they might also look similar when they both grow older. By leveraging a large-scale image dataset freely available on the Internet as a reference set, CARC is able to encode the low-level feature of a face image with an age-invariant reference space. Two images of the same person have similar representations using CARC for the reason that they both look similar to certain reference people with different ages.

Recently, deep learning has received much attention in the research field of machine learning due to its superior performance in learning a series of nonlinear feature mapping functions directly from raw features. It has been widely used in many computer vision topics and obtain very promising results. Some deep learning methods have also been proposed for age-invariant face recognition to improve the performance [1, 13, 17, 19, 25, 29]. Zhai et al. [29] combine deep convolutional neural network and local binary pattern histograms to improve the recognition accuracy. Lu et al. [19] propose a new joint feature learning (JFL) and stack this model into deep architecture for face representation. Lin et al. [17] incorporate the similarity measure matrix into the deep architecture, enabling an end-to-end way of model optimization. They unify the simi-
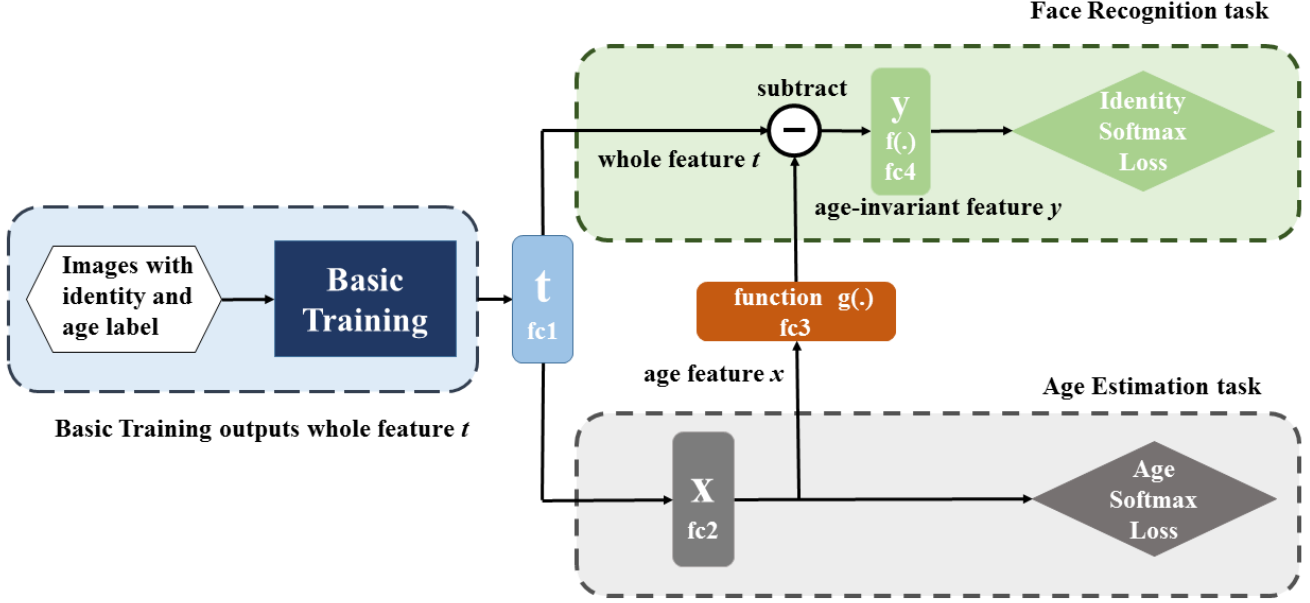
Figure 2. The architecture of the proposed AE-CNN. The formulation we use is $y = f(t - g(x))$ as shown in (1), t is the whole feature which contains age-related factor, x is the age feature obtained in age estimation task, y is the identity-specific feature for age-invariant face recognition, $g(.)$ is the function to obtain age factor which degrades the performance of face recognition from age feature. $f(.)$ is the function to better handle the relationship between the whole feature, age feature and identity-specific feature. The age estimation task and the face recognition task update parameters in the network at the same time.

larity measure with feature representation learning via deep convolutional neural networks. Wen *et al.* [25] propose a model called latent factor guided convolutional neural networks (LF-CNNs) to address the age-invariant face recognition task. They develop a latent variable model called latent identity analysis (LIA) and by coupled learning the parameters in CNNs and LIA, the age-invariant deep features can be extracted. Xu *et al.* [27] propose a nonlinear method to separate whole feature to get identity feature and present a neural network model called coupled auto-encoder networks (CAN) which leverages two shallow neural networks as bridge connecting the two auto-encoders to fit aging and de-aging process.

3. The Proposed Method

In this section, we describe the proposed method. We first introduce formulations to obtain age-invariant feature by using age feature and the whole feature, and then present the deep architecture and training algorithm for the proposed AE-CNN.

3.1. Model Formulation

Matching cross-age face images is necessary in real world and the difficulty is that aging process causes large intra-class variations including shape changes and texture changes. Through observation we can find that face images of different people in the same age usually share character-

istics in common such as wrinkles and skin. On the other hand, face images of the same person also keep features that are relatively stable across ages such as gender.

The person-specific feature is what we need in age-invariant face recognition problem. However, the person-specific feature is latent and getting it directly is difficult since the feature we obtain by common face recognition task often contains age-related component. Considering the fact that the age feature can be easily acquired through age estimation task, we propose a new method using age estimation task to guide face recognition. We use age feature to get age factor which degrades the performance of face recognition and obtain age-invariant feature by removing age factor from the whole feature. Specifically, we use vectors to represent the age feature, identity-specific feature and the whole feature of the input image. The formulation expresses that the identity-specific feature can be obtained by using the whole feature and the age feature. Overall, the formulation can be written as:

$$y = f(t - g(x)), \tag{1}$$

$$f(x) = W_1 x + b_1, \tag{2}$$

$$g(x) = W_2 x + b_2, \tag{3}$$

t is a $d \times 1$ vector representing the whole face feature, x is a $d \times 1$ vector representing the age feature obtained by age estimation, y is a $d \times 1$ vector representing the latent

identity-specific feature which we need, $g(.)$ is the function to obtain age factor which degrades the performance of face recognition from age feature. $f(.)$ is the function to better handle the relationship between the whole feature, age feature and identity-specific feature.

The basic idea of our approach is to obtain the latent identity-related feature which is difficult to get directly by using the whole feature of the image which can be got from the first fully connected layer and the age-related factor which can be acquired through age estimation task.

3.2. The AE-CNN Framework

An overview of the proposed CNN framework is shown in Figure 2.

Structurally the AE-CNN is composed of convolution layers and fully-connected layers as functions in (1) to get age-invariant feature. The structure of the convolutional component in the proposed framework follows typical CNNs, alternatively stacking convolution layer, nonlinear layer and max-pooling layer. We use the CNN structure in [26] called lightened CNN which is constructed with 4 convolution layers as shown in Figure 3. The convolution kernel sizes are set as 9×9, 5×5, 5×5, 4×4 and the stride is set as 1. To improve model fitting, we use Max-Feature-Map (MFM) activation function. The max pooling layers are used to enhance robustness to potential translation and sub-sampling.

Then we focus on the construction of the fully-connected layers to separate latent age-invariant feature from age-specific factor. The fully-connected layer fc1 outputs the whole feature of the facial image, fc2 outputs the age feature. Notice that the fully-connected layer is equivalent to matric multiplication: $F_{fc} = W F_{in} + b$, where W, b are the parameters of the fully-connected layer, fc3 represents the functions in (1) as $g(.)$ and the eltwise layer subtracts the age factor from the whole feature. Finally, fc4 represents the functions in (1) as $f(.)$ and outputs the identity-specific feature as shown in Figure 2.

The inputs of the proposed CNN framework are training face images with identity labels and age labels. For the input image, our goal is to separate person-specific feature from age factor by subtracting age-specific component from the whole feature. To obtain age-invariant feature, in our model, specifically we have two steps:

1. **Basic Training**: This step achieves face recognition using the framework without age estimation and outputs fc1 in Figure 2 as features. The cost function is defined as:

$$L_1 = -log(\frac{e^{t_i^p}}{\sum_{j=1}^m e^{t_i^j}}), \quad (4)$$

where t_i^p is the identity classifier output in class p of the i th input. The feature we obtain still contains age-related

factor, it is the whole representation of the input facial image.

2. **Separation**: This step achieves age estimation task and face recognition task at the same time and outputs fc4 in Figure 2 as features. The cost function is defined as:

$$L_2 = -log(\frac{e^{t_i^p}}{\sum_{j=1}^m e^{t_i^j}}) - \alpha(log(\frac{e^{a_i^r}}{\sum_{k=1}^n e^{a_i^k}})), \quad (5)$$

where t_i^p is the identity classifier output in class p of the i th input and m is the number of identity classes, a_i^r is the age classifier output in class r of the i th input and n is the number of age classes. The first component of L_2 is used for face recognition and the second component is used for age estimation. α is the loss weight of age estimation task. We change it during the training process to change the importance of age estimation task to get better results. This step uses age estimation to guide age-invariant face recognition, the two tasks update the parameters in network simultaneously. We use the model got in step 1 to initialize the network, add fully-connected layers as functions to better handle the relationship between age feature, identity-specific feature and the whole feature to obtain better person-specific feature. By applying age estimation task we can obtain age-specific factor which degrades the performance of age-invariant face recognition. Then we remove it from the whole feature and the person-specific feature can be obtained.

The two steps above build the proposed AE-CNN.

3.3. Training

Training the proposed AE-CNN involves two steps as discussed above and we describe our training procedure in Algorithm 1. To solve (5), we adopt stochastic gradient descent (SGD) using standard back propagation to update parameters we need to obtain person-specific feature for recognition.

4. Experiments

In this section, we evaluate our approach on two public aging face datasets: MORPH Album 2 [23] and Cross-Age Celebrity (CACD) [2] to demonstrate the effectiveness of the proposed method.

4.1. Implementation Details

Datasets. MORPH Album 2 [23] contains more than 55,000 face images of more than 13,000 individuals. Age ranges from 16 to 77 and the average number of images per person is 4. CACD [2] is a large scale dataset released in 2014. The dataset contains more than 160,000 images
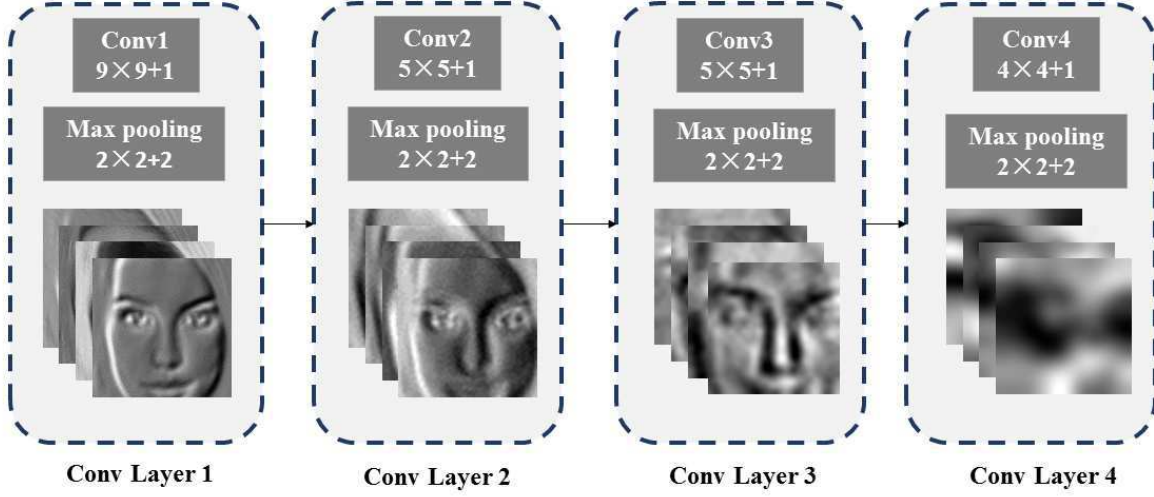
Figure 3. The architecture of convolutional component.

Algorithm 1 Learning Algorithm for AE-CNN

Input: Cross-age training data y_i with both age label r and identity label p, testing image y_i'

Output: The age-invariant feature

1: **The learning in basic training step:**
   Train basic network with training data y_i and identity label p, use softmax loss function in (4) to obtain the whole feature t which includes age-related factor.

2: **The learning in separation step:**
   (1) Initialize the parameters using model got in basic training step.
   (2) Train separation network with training data y_i, identity label p and age label r, use loss function in (5) to obtain age-invariant feature y for cross-age face recognition.

3: **For each testing image:**
   Forward y_i' to the trained CNN and get age-invariant feature for testing.
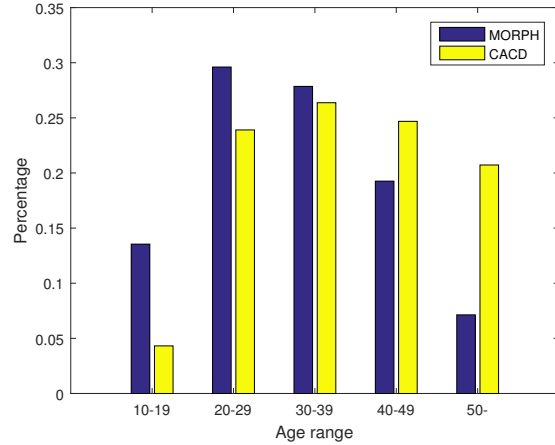

Figure 4. Age range distribution of MORPH Album 2 and CACD.

Figure 5. Face images alignment for MORPH Album 2 dataset. (a) is the facial detection result and (b) is the normalized face image.

of 2,000 celebrities with age ranging from 16 to 62. The images are collected from the Internet. Figure 4 shows the age distribution of these two datasets.

Preprocessing. It is necessary to preprocess the image database. Facial landmarks are localized by AAM [4] and then adjusted by hand, all the faces are aligned by the landmarks of eyes and the midpoint of mouth corners. After that, pictures with the size of 144×144 are cropped from the aligned images. To avoid overfitting when training the proposed AE-CNN, we do data augmentation. Images with the size of 144×144 are randomly cropped to 128×128 before being fed to the convolution layers. All the input

images are gray-scale face images as shown in Figure 5.

Detailed parameters setting. Our model is finished by caffe library [10]. The batch size is 64. When training the basic network which only achieves face recognition and
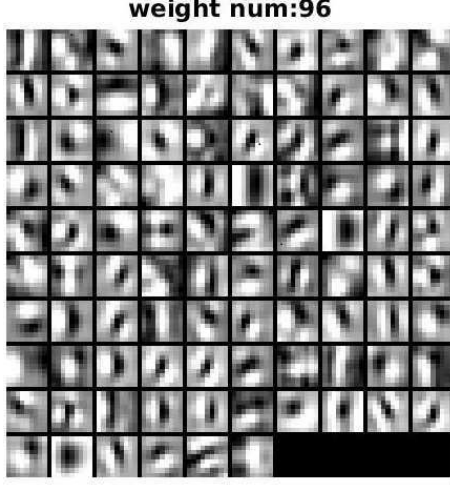
Figure 6. Learned convolutional kernels in the first convolution layer of AE-CNN.

Figure 7. The performance of MAP of the proposed method compared with state-of-the-art algorithms on CACD.

the proposed AE-CNN, the learning rate is 1e-3, 1e-4 and is switched when the loss plateaus. The loss weight α of age estimation in (5) is 0.5 for MORPH Album 2 and 1 for CACD training dataset.

Classifier. To evaluate the performance of the proposed person-specific feature, our approach uses cosine distance and the nearest neighbor rule as the classifier.

4.2. Experiments on CACD Dataset

CACD dataset contains images of 2,000 celebrities across ten years with age ranging from 16 to 62. The images in this dataset have vary illumination, different poses, different make up and better simulate practical scenario.

We follow the experimental setting in [2,27], choose 120 celebrities with rank 3-5 as test sets. The images taken at 2013 are used as query images and the remaining images are split into three subsets and are used as gallery images. The first subset contains images taken in 2004-2006, the second subset contains images taken in 2007-2009 and the third contains images taken in 2010-2012. During the train process, the rest of images in CACD are used as training data with identity label and age label to update parameters.

We plot the learned kernels in the first convolution layer of the proposed AE-CNN in Figure 6. These kernels are convoluted with the input image to extract discriminative features. It can be seen that most kernels are simply edges and spots, which can effectively extract discriminative information on the human face.

In our experiment on CACD, we use mean average precision (MAP) as evaluation protocol. Cosine distance is used to compute the similarity of two images. Specifically, let $q_i \in Q$ be the query images and Q is the query image dataset. For q_i, the number of relevant images is m_i and
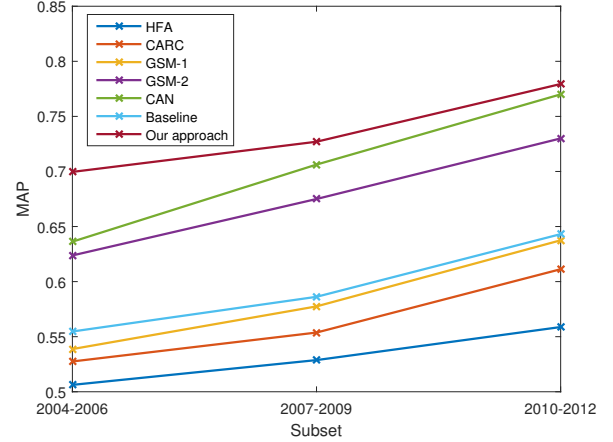
the relevant images can be expressed as $Y_1, Y_2, \ldots, Y_{m_i}$. E_{ic} is the retrieval results of q_i in a descending order from the top to Y_c. And the average precision (AP) of q_i can be calculated as below:

$$AP(q_i) = \frac{1}{m_i} \sum_{c=1}^{m_i} Precision(E_{ic}), \qquad (6)$$

where $Precision(E_{ic})$ means the ratio of relevant images in E_{ic}. Then the average of the AP of all images (MAP) of Q can be calculated as:

$$MAP(Q) = \frac{1}{|Q|} \sum_{i=1}^{|Q|} AP(q_i). \qquad (7)$$

Using this protocol, we compare our method with the existing methods in this dataset including hidden factor analysis (HFA) [8], cross-age reference coding (CARC) [3], generalized similarity model [17] (GSM-1 and GSM-2, compared with GSM-1, GSM-2 only uses more training data) and coupled auto-encoder networks (CAN) [27]. Figure 7 reports the comparative results. All methods in Figure 7 are tuned to the best setting according to their papers. For fair comparison, we also compare the proposed AE-CNN with a baseline CNN model. The baseline CNN is the network we use in training process step 1 which has the same convolution unit and first fully connected layer in the proposed method but only achieve face recognition task. The baseline outputs the whole feature t which contains age-related factor.

From the results in Figure 7, we have the following observations. First, our method outperforms the others in all three subsets. Note that compared with other methods which perform well with small age gaps, our method achieves competitive performance with large age gaps. The

Figure 8. Some examples of failed retrievals in CACD dataset. The first row is the probe faces, the second row presents the incorrect rank-1 matching results using the proposed approach and the bottom row shows the corresponding gallery images for the probe images.

Table 1. Recognition rates on MORPH.

Method	Rank-1 Identification Rates
HFA (2013) [8]	91.14%
CARC (2014) [3]	92.80%
MEFA (2015) [9]	93.80%
MEFA+SIFT+MLBP (2015) [9]	94.59%
LPS+HFA (2016) [15]	94.87%
GSM (2016) [17]	94.40%
LF-CNNs baseline (2016) [25]	95.13%
LF-CNNs (2016) [25]	97.51%
CNN baseline (trained by CASIA data)	74.73%
CNN baseline (fine-tuned by MORPH data)	96.30%
AE-CNN	**98.13**%

proposed AE-CNN improves the MAP from 0.64 to 0.70 in 2004-2006 subset. This confirms the superiority of our approach. Second, it is encouraging to see that the proposed AE-CNN obtains a significant improvements over the CNN-baseline results. The improvement shows that the whole feature we obtain by face recognition task contains age-related factor and degrades the performance of age-invariant face recognition. However, by adding age estimation and removing the age factor from the whole feature, we successfully obtain person-specific feature, and the person-specific feature is very powerful in cross-age face recognition.

Figure 8 shows some examples of the failed retrievals. While the rank-1 retrievals are not correct in these cases, the probe images appear to be more similar to the incorrect retrievals than the true images.

4.3. Experiments on MORPH Album 2 Dataset

MORPH Album 2 [23] consists of more than 55,000 face images of more than 13,000 individuals with age ranging from 16 to 77. 10,000 individuals are used for training and the remaining 3,000 individuals are used for testing, there is no overlapping subject between the training set and the testing set. The testing set is composed of a probe set and a gallery set, two face images with the youngest age and the oldest age are selected as gallery and probe set respectively for each subject. This experimental setting is the same with those adopted in [2, 8, 17]. Considering that the number of images in training set is not very big, we use the model trained on CASIA-WebFace [28] and fine-tune for face recognition to avoid overfitting.

We compare the proposed CNN model with (1) the CNN baseline model which is trained by CASIA-WebFace using only face recognition. (2) the CNN baseline model which is

fine-tuned by MORPH training set. (3) some state-of-the-art approaches on the dataset. The comparative results are reported in Table 1.

From Table 1 we have the following conclusions. First, the result of baseline CNN without fine-tuning is only 74.73%. It is inferior to the other results in Table 1. This confirms that the parameters trained using CASIA-WebFace dataset [28] are not suitable for MORPH Album 2 [23] which shows that images in CASIA dataset and images in MOPRH Album 2 are different and directly applying the CNN model to address age-invariant face recognition is not a good choice. Second, the CNN baseline model fine-tuned by MORPH Album 2 can reach an accuracy of 96.3%, which shows that the convolutional neural network structure is powerful. However, the features obtained by baseline model cannot remove age-related factor by itself, thus the performance of cross-age face recognition degrades. The accuracy reflects the limitation of common CNN model which only uses face recognition loss function, so it is desirable to design a new deep CNN model which can remove age factor from images and can get person-specific feature for cross-age face recognition. The proposed AE-CNN achieves comparative result of 98.13% on the MORPH Album 2 dataset. This confirms the effectiveness of our algorithm. Moreover, the proposed method outperforms the CNN baseline by a clear margin. This shows that using age estimation task to guide age-invariant face recognition task is useful. By subtracting age factor from the whole feature, we successfully obtain person-specific feature for age-invariant face recognition.

In MORPH Album 2, age distribution of the subjects is not uniform, so it is desirable to exploit the influence of different age ranges on the proposed method. We separate the age range into five age groups by 16-19, 20-29, 30-39, 40-

Table 2. Performance of different age groups on MORPH.

Age group	Amount	CNN baseline	AE-CNN
16-19	337	97.03%	**97.97%**
20-29	940	95.64%	**97.66%**
30-39	779	95.76%	**97.69%**
40-49	695	96.98%	**98.99%**
50-77	249	97.99%	**99.20%**

Table 3. Performance of different measurements on MORPH 2.

| Measurement	CNN baseline	AE-CNN
Mahalanobis distance	84.74%	**87.60%**
Euclidean distance	94.90%	**96.47%**
Cosine distance	96.33%	**98.13%**

Figure 9. The performance of age estimation and cross-age face recognition in every 1000 iterations.
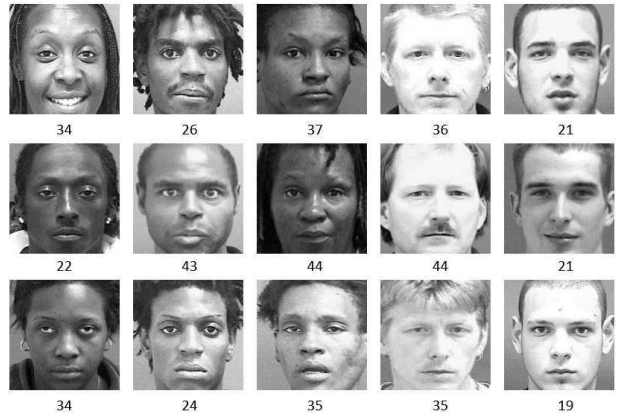

Figure 10. Some examples of failed retrievals in MORPH Album 2 dataset. The first row is the probe faces, the second row is the incorrect rank-1 matching results using the proposed approach, and the bottom row shows the corresponding gallery images for the probe images.

49 and 50-77. Table 2 shows the rank-1 identification rates in different age groups. According to Table 2, we can see that the proposed approach outperforms the baseline CNN model on all age groups, which shows that our method can obtain person-specific feature which can perform well in age-invariant problem. The results further confirms the advantage of our method over common CNN model in cross-age face recognition problem.

Experiments are also performed to measure the similarity metric methods. We use Euclidean distance, Mahalanobis distance and cosine distance. According to the results shown in Table 3, cosine distance outperforms the other measurements.

To further evaluate the effect of age estimation, we design an experiment to report the age estimation results and face recognition results in every 1,000 iterations during the training process. The parameters to separate person-specific feature are updated during the training process. We use rank-1 identification rate to estimate the performance of cross-age face recognition and we utilize the mean absolutely error (MAE) to measure the error between the predicted age and the ground-truth, which is normalized and defined as follows:

$$MAE = \frac{\|y' - y\|_2}{N},$$ (8)

where y' and y denote predicted age value and ground-truth age value. N denotes the number of the testing samples. Figure 9 clearly shows that the MAE of age estimation and the accuracy of face recognition improve at the same time, which shows that age estimation task guides age-invariant face recognition.

Finally, we show some failed retrievals in MORPH dataset in Figure 10. The results confirm what we observed from Figure 8: The rank-1 retrieved images appear highly similar to the probe images in the incorrect matching.
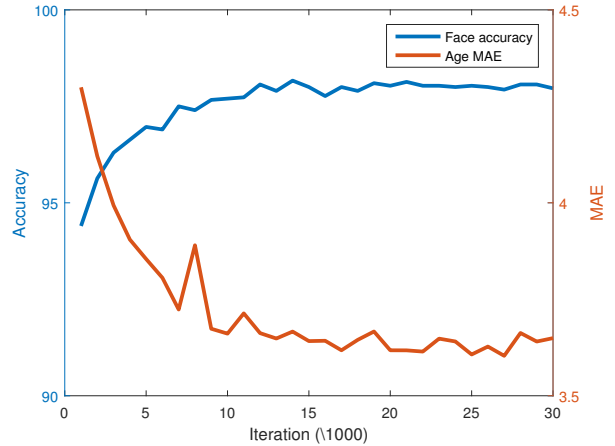
5. Conclusions

In this paper, we have proposed an age estimation guided CNN approach to address the challenging problem of age-invariant face recognition. Unlike the existing deep learning models, we use age estimation task. Considering the fact that directly obtaining person-specific feature is difficult since the feature we get by face recognition task always contains age-related factor, we add age estimation task to obtain age feature and subtract age factor from the whole feature. Extensive experiments have been conducted on two public domain face aging datasets (CACD and MORPH Album 2) to confirm the effectiveness of our approach.

References

[1] S. Bianco. Large age-gap face verification by feature injection in deep networks. *CoRR*, abs/1602.06149, 2016. 2

[2] B.-C. Chen, C.-S. Chen, and W. H. Hsu. *Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval*, pages 768–783. Springer International Publishing, Cham, 2014. 2, 4, 6, 7

[3] B. C. Chen, C. S. Chen, and W. H. Hsu. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Transactions on Multimedia*, 17(6):804–815, June 2015. 6, 7

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, Jun 2001. 5

[5] J.-X. Du, C.-M. Zhai, and Y.-Q. Ye. *Face Aging Simulation Based on NMF Algorithm with Sparseness Constraints*, pages 516–522. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 1, 2

[6] C. Geng and X. Jiang. Face recognition using sift features. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3313–3316, Nov 2009. 2

[7] X. Geng, Z. H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2234–2240, Dec 2007. 2

[8] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden factor analysis for age invariant face recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013. 1, 2, 6, 7

[9] D. Gong, Z. Li, D. Tao, J. Liu, and X. Li. A maximum entropy feature descriptor for age invariant face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 1, 2, 7

[10] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 675–678, New York, NY, USA, 2014. ACM. 5

[11] B. Klare and A. K. Jain. Face recognition across time lapse: On learning feature subspaces. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–8, Oct 2011. 2

[12] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, Apr 2002. 2

[13] Y. Li, G. Wang, L. Lin, and H. Chang. *A Deep Joint Learning Approach for Age Invariant Face Verification*, pages 296–305. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. 2

[14] Z. Li, D. Gong, X. Li, and D. Tao. Learning compact feature descriptor and adaptive matching framework for face recognition. *IEEE Transactions on Image Processing*, 24(9):2736–2745, Sept 2015. 2

[15] Z. Li, D. Gong, X. Li, and D. Tao. Aging face recognition: A hierarchical learning model based on local patterns selection.

[16] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE Transactions on Information Forensics and Security*, 6(3):1028–1037, Sept 2011. 1, 2

[17] L. Lin, G. Wang, W. Zuo, F. Xiangchu, and L. Zhang. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2016. 2, 6, 7

[18] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. Face verification across age progression using discriminative methods. *IEEE Transactions on Information Forensics and Security*, 5(1):82–91, March 2010. 1, 2

[19] J. Lu, V. E. Liong, G. Wang, and P. Moulin. Joint feature learning for face recognition. *IEEE Transactions on Information Forensics and Security*, 10(7):1371–1383, July 2015. 2

[20] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, Jul 2002. 2

[21] C. Otto, H. Han, and A. Jain. *How Does Aging Affect Facial Components?*, pages 189–198. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. 2

[22] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):947–954, May 2010. 1

[23] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345, April 2006. 2, 4, 7

[24] X. Wang and X. Tang. Random sampling lda for face recognition. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–259–II–265 Vol.2, June 2004. 2

[25] Y. Wen, Z. Li, and Y. Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2, 3, 7

[26] X. Wu, R. He, and Z. Sun. A lightened cnn for deep face representation. In *2015 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, 2015. 4

[27] C. Xu, Q. Liu, and M. Ye. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing*, 222:62 – 71, 2017. 3, 6

[28] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014. 7

[29] H. Zhai, C. Liu, H. Dong, Y. Ji, Y. Guo, and S. Gong. *Face Verification Across Aging Based on Deep Convolutional Networks and Local Binary Patterns*, pages 341–350. Springer International Publishing, Cham, 2015. 2

IEEE Transactions on Image Processing, 25(5):2146–2154, May 2016. 1, 2, 7