Pattern Recognition Letters 125 (2019) 271–277

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Recurrent age estimation☆

Huiying Zhang[a,b,c,*], Xin Geng[b], Yu Zhang[b], Fanyong Cheng[d]

[a] Pujiang Institute, Nanjing Tech University, Nanjing 211200, China
[b] School of Computer Science and Engineering, Southeast University, Nanjing 211189, China
[c] Department of Computer Engineering, Bengbu College, Bengbu 233030, China
[d] College of Electrical Engineering, Anhui Polytechnic University 241000, China

ARTICLE INFO

Article history:
Received 24 January 2019
Available online 2 May 2019

Keywords:
Recurrent age estimation (RAE)
Convolutional neural network (CNN)
Long short-term memory (LSTM)
Age estimation
Label distribution learning (LDL)

ABSTRACT

Age estimation is a challenging research topic in recent years. Existing approaches usually use only appearance features for age estimation. Personalized aging patterns, i.e., sequences of personal features, which have been shown as an important factor for improving age estimation accuracy, however, are not considered in their researches. We propose a novel model named recurrent age estimation (RAE), to make full use of appearance features as well as personalized aging patterns. RAE uses the CNN-LSTM architecture. Convolutional neural networks (CNNs) are trained to extract discriminative appearance features from face images, and long short-term memory networks (LSTMs) are employed to learn personalized aging patterns from sequences of personal features. Furthermore, we integrate the label distribution learning (LDL) scheme into LSTMs to exploit ambiguity from the real age and adjacent ages. The superiority of the RAE compared with existing approaches is shown by experimental results.

1. Introduction

Age is an important biometric trait of human beings, and it plays a fundamental role in social interaction. Many researchers have been devoted to age estimation as it can be applied in a variety of fields [3,28], e.g., enhancing the human-computer interaction, preventing children from buying alcohol, tobacco and visiting adult websites, finding the lost children and analyzing consumer ages.

In recent years, many age estimation methods have been proposed, most of which belong to one of the following three fundamental categories: hand-crafted methods [11,30], shallow learning [10,19,25] and deep learning methods [8,21]. Hand-crafted methods extract features from appearance, shape, facial texture and skin color, which require prior and expert knowledge. Shallow learning extract features by local binary methods from facial raw pixels of local face patches. In contrast, deep learning methods automatically learn effective facial features through CNNs. Many experimental results have shown that deep learning methods achieve better results than hand-crafted methods and shallow learning methods [4,9].

The performance of deep learning methods depend on a large labeled training set. When lacking of labeled face images, deep learning methods tend to be over-fitting on small image collections. Unfortunately, existing databases do not contain sufficient labeled face images, as it is expensive to collect people photos in their whole lifetime. To overcome this problem, Geng et al. [13] proposed LDL which exploits ambiguity from the real age and adjacent ages. They applied LDL with hand-crafted method to age estimation and improved age prediction accuracy [14,16]. Subsequently, Gao et al. [12] applied LDL in visual representations with convolutional networks named deep label distribution learning (DLDL). The experimental results have shown that DLDL is superior to existing methods. The two approaches deal with face images separately and focus on static appearance features, but ignore aging patterns.

Aging patterns are series of individual face images [17]. People deduce others' ages based on not only their current features but also evolutional information of features. Therefore, we take account of this factor in our RAE model in addition to people's appearance features. It is important for us to note that aging pattern is uncontrollable and complicated, and it varies for different individuals due to the influence of genes, ethnicity, gender, living habits, health status, living environment, and so on. As shown in Fig. 1, three people are very different in appearance change. Existing databases, however, have a limited number of face images, which makes it difficult to learn people's aging patterns.

☆ Conflict of interest: None.

* Corresponding author.
E-mail address: bbzhy@126.com (H. Zhang).

https://doi.org/10.1016/j.patrec.2019.05.002
0167-8655/

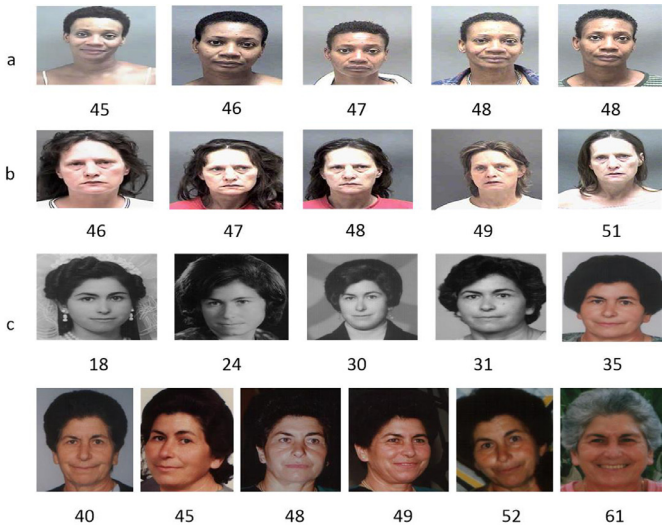


Fig. 1. Aging faces of three individuals, (a) and (b) in the Morph, and (c) in the FG-NET. The numbers under the pictures are their corresponding real ages.

Experimental results have shown that recurrent neural networks (RNNs) yield superior performance in sequential data [1]. As an extension of RNN, LSTM is able to effectively learn long-term dependency information through gates structure [7]. So we employ LSTM to learn personalized aging patterns.

We train CNNs to extract facial image features and feed the features into LSTMs to learn aging patterns. Moreover, we introduce a more practical LDL into LSTMs to exploit label ambiguity, thus overcoming the problem of small data sets.

The contributions of this paper are:

(1) Compared with existing approaches, the proposed RAE model utilizes LSTM for leaning personalized aging patterns improve the prediction of age.

(2) A more practical LDL is proposed. Given any age label, we will limit the LDL that only covers a fixed and reasonable adjacent ages. LDL expands the number of face images.

(3) RAE is flexible enough to handle the sequence of images with different lengths. We test RAE on two public databases, and the results show that RAE is superior to existing approaches.

The paper is organized in this way. The related work is review in Section 2. The relevant concepts including time series, aging pattern, LDL, improved LDL and the proposed RAE are introduce in Section 3. Our experiments and results are introduce in Section 4. Finally, our research is summarized in Section 5.

2. Related work

Age estimation is a challenging research topic because aging is complicated and uncontrollable. It is difficult to find a robust and accurate algorithm for age estimation. Automatic age estimation based on facial structure is the earliest age estimation technique developed in the past 20 years.

Age estimation is divided into two phases: (1) feature extraction from face images; (2) regression or classification based on features. Many age-related feature extraction algorithms have been proposed, such as anthropometric model [22], active appearance model [5], active shape model [30], and aging manifold model [11]. These approaches mainly rely on either shape or texture features that are extracted from face images. To improve accuracy of age estimation, some models based on the combination of these features are proposed [6]. However, experimental results show that there still exists a big gap between the real and the estimated age.

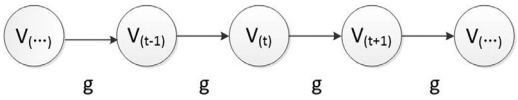


Fig. 2. Unfold dynamic computation graph.

Geng et al. [17] proposed AGES with hand-crafted method. The idea is to establish aging subspaces for each individual in chronological order, and then the position of the face image is estimated age. Biologically inspired features (BIF) [20,32] employs some predefined Gabor filters to learn facial features. Gabor filter is good at capturing edge information, but not suitable for the complicate facial feature. Lu et al.[26] proposed cost-sensitive local binary feature learning (CS-LBFL) approach to learn features from facial raw pixel.

Deep learning networks, especially CNNs and RNNs have shown tremendous progress in many pattern recognition fields [7,31] such as face detection, emotion recognition, speech recognition, and so on. CNNs extract more obviously discriminative features. RNNs process sequential data.

Liu et al. [24] presented a label-sensitive deep metric learning (LSDML) method for facial age estimation by deep residual network. Recently, Liu et al. [23] proposed an ordinal deep feature learning (ODFL) method to learn ordinal features by deep Convnet. They further proposed ordinal deep learning (ODL) to complement information of ODFL. Some extended CNNs are proposed to improve the accuracy of age estimation [4,9]. Although these methods made some progress, the gap between the real and the estimated age is still large. In addition, researches show that estimation accuracy is difficult to improve by adding network depth [2]. This is because the facial image information is difficult to capture.

3. Proposed approach

In this section, the definition of time series is first introduced. Then, we describe LDL and improved LDL. On this basis, we introduce RAE.

3.1. Time series and aging pattern

Time series (dynamic data) are sequential data collected at different time steps. These data reflect the inherent patterns of certain things. Consider the classical form of a dynamic system [18]:

$$v_{(t)} = g(v_{(t-1)}; \boldsymbol{\theta}), \tag{1}$$

where $v_{(t)}$ is the data at time t. The prediction of $v_{(t)}$ depends on the sequence before time t. Each data $v_{(t)}$ can be calculated by the same definition and uses the same parameter vector $\boldsymbol{\theta}$, so (1) is recurrent function. For a more intuitive understanding, the unfolded computation graph of (1) is illustrated in Fig. 2.

Aging pattern is accord with sequential data. The age sequence does not have a fixed length, t is a variable. Let vector $\mathbf{x}_t$ indicate the image at time t. We consider an aging system driven by face image $\mathbf{x}_t$:

$$v_{(t)} = g(v_{(t-1)}, \mathbf{x}_t; \boldsymbol{\theta}), \tag{2}$$

The recurrent function g learns the aging pattern from sequential personal images.

Let $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)$ be the sequence of personal images, and $Y = (y_1, y_2, \ldots, y_t)$ be the age, corresponding to the sequence of images. In order to learn the aging pattern, the whole sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)$ inputs (2). The purpose of age estimation is to learn the parameter vector $\boldsymbol{\theta}$, so that the age y_t of image $\mathbf{x}_t$ can be calculated with the given sequence of images.

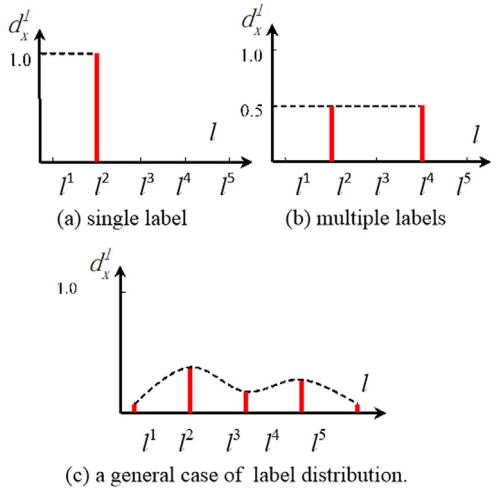


Fig. 3. Three forms of label distribution.

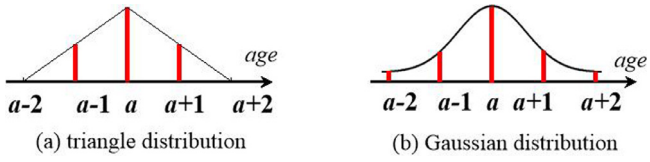


Fig. 4. Two label distributions at age a.

3.2. LDL and improved LDL

3.2.1. LDL

Label distribution learning (LDL) was first proposed in 2014 [13] and has attracted attention in machine learning recently [15,33]. LDL is used to reflect the importance of different labels in learning algorithm.

For an instance $\mathbf{x}$, $d_{\mathbf{x}}^{l} \in [0, 1]$ indicates description degree which is label l describes $\mathbf{x}$. Fig. 3 shows three forms of label distribution for five class labels. For case (a), there only one label l_2 describes $\mathbf{x}$, so $d_{\mathbf{x}}^{l_2} = 1$. For case (b), two labels l_2 and l_4 describe $\mathbf{x}$, the two description degrees $d_{\mathbf{x}}^{l_2}$ and $d_{\mathbf{x}}^{l_4}$ are both 0.5. (c) is a universal label distribution that satisfies $d_{\mathbf{x}}^{l_i} \in [0, 1]$ and $\sum_i d_{\mathbf{x}}^{l_i} = 1$. Single label and multiple labels are the special forms of label distribution.

In age estimation, adjacent faces are similar to each other. Age can be labeled with adjacent ages. For example, when a person's actual age is 20, we say that the person is "about20" years old, which covers a reasonable number of neighboring ages. Not only the real 20 but also 19(21) and 18(22) can be used to describe the age, 19(21) is closer than 18(22). Facial age can be converted to age label distribution.

It is reasonable to set a person's age between 1 and 100. $\ell = \{l_1, l_2, \ldots, l_{100}\}$ denotes all labels corresponding to face image $\mathbf{x}$. The description degree of l to nth face image $\mathbf{x}_n$ is $d_{\mathbf{x}_n}^{l}$. Let $\mathbf{D}_{\mathbf{x}_n} = \{d_{\mathbf{x}_n}^{l_1}, d_{\mathbf{x}_n}^{l_2}, \ldots, d_{\mathbf{x}_n}^{l_{85}}\}$ indicate the label distribution of $\mathbf{x}_n$,

$T = \{(\mathbf{x}_1, \mathbf{D}_{\mathbf{x}_1}), (\mathbf{x}_2, \mathbf{D}_{\mathbf{x}_2}), \ldots, (\mathbf{x}_N, \mathbf{D}_{\mathbf{x}_N})\}$. The purpose of the age LDL is to learn a condition probability distribution function $\mathbf{D}^* = p(\mathbf{D}|\mathbf{x}; \boldsymbol{\theta})$ from T. $\mathbf{D}$ is the real distribution and $\mathbf{D}^*$ is the predicted distribution. The fine parameter vector $\boldsymbol{\theta}$ needs to be learned so that $\mathbf{D}^*$ is similar to $\mathbf{D}$.

3.2.2. Improved LDL

Fig. 4 shows two label distributions of the real age a, in which one is triangle distribution, and the other is Gaussian distribution. We can see that label distribution is a continuous function, as the age is a discrete and integer number. The probability distribution of continuous variables is converted to description degree.

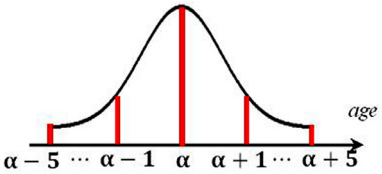


Fig. 5. Improved label distribution at the real age based on Gaussian distribution.

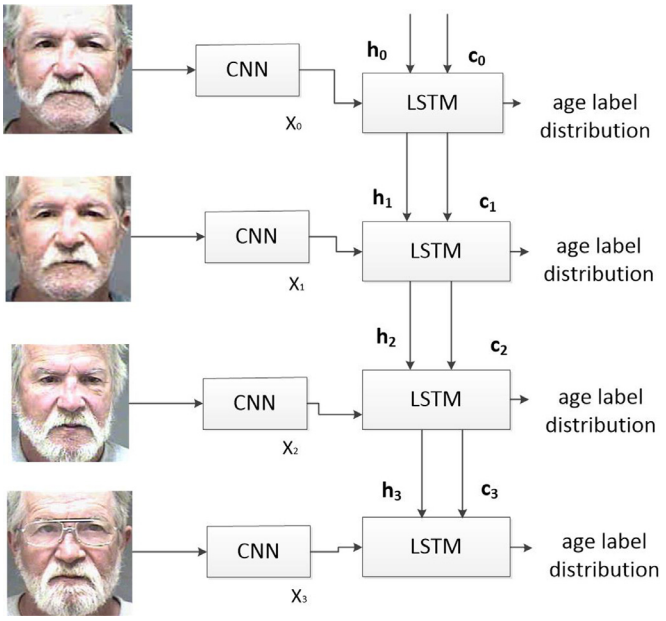


Fig. 6. The pipeline of RAE.

According to the literature [16], Gaussian distribution is superior to triangular distribution. We improve Gaussian distribution to simulate the label distribution of facial image age. The age label distribution $\mathbf{D}_{\mathbf{x}_n}$ should satisfy with two basic conditions: (1) The probability distribution should be based on the real age. The probability of the real age is the biggest, and the greater the distance away from real age, the smaller the probability is. (2) $\sum_m d_{\mathbf{x}}^{l_m} = 1$ and $d_{\mathbf{x}}^{l_m} \in [0, 1]$. According to the Gaussian distribution function, the real age is a, and the function is

$$p(l_m|\mu; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(l_m - a)^2}{2\sigma^2}\right). \tag{3}$$

Suppose we have a person's 20-year-old face image. It is impossible for us to learn correlate information from the person's 60-year-old face image. Therefore, we consider it unacceptable that the gap between a person's real age a and the predicted age is greater than five. Inspired by this idea, we consider a simplified age label distribution shown in Fig. 5, which covers a fixed range of ages $[a - 5, a + 5]$. In the range, each label is assigned with a description degree to describe the image. The description degree is zero when out of the range.

3.3. RAE

The purpose of RAE is to learn personalized aging patterns to predict the age. RAE is illustrated in Fig. 6. Our pipeline consists of two distinct deep learning frameworks: CNN and LSTM. We train CNNs to extract image features. Next, personal sequential features are fed into LSTMs directly. LSTMs simultaneously learn the per-

sonalized aging pattern and predict ages. In LSTMs, each face has a label distribution by introducing LDL.

3.3.1. Image feature extraction

Given a CNN model, our goal is to train a fine parameter vector that is adapted to the database to extract the discriminative features.

$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$, y_i is the label of the face image $\mathbf{x}_i$. The goal of age estimation is to learn a condition probability function $y^* = p(y|\mathbf{x}; \boldsymbol{\theta})$, in which y is the given label, and y^* is the predicted age. The best $\boldsymbol{\theta}$ is the training goal of CNN, which minimizes the distance between y and y^*.

The softmax layer assigns a probability for each age classification. The total loss of cross entropy (4) is employ to optimize the CNN model.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} \mathbf{P}_{\mathbf{x}_i} \log \mathbf{Q}_{\mathbf{x}_i}, \tag{4}$$

here, the probability distribution $\mathbf{P}_{\mathbf{x}_i}$ is the expected output of face image $\mathbf{x}_i$, the probability distribution $\mathbf{Q}_{\mathbf{x}_i}$ is the output of softmax layer. Finally, the parameter vector $\boldsymbol{\theta}$ of the trained model is saved and used to extract effective image features by dropping the softmax layer.

Based on CNNs, we can only extract face image features, and cannot learn dependency information from sequential face images. Everyone has a unique aging pattern, so it is unreasonable to judge a person's age only from facial features. Therefore, we introduce LSTMs to learn personalized aging patterns.

3.3.2. LSTM-based model

LSTMs can effectively handle sequential data of arbitrary lengths and solve gradient explode and vanish. The LSTMs model is trained to learn aging pattern from sequence of individual features and to predict ages.

Let $\mathbf{x}$ represent the features extracted from CNNs. Given a sequence of individual features $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t)$ and the corresponding ages $\mathbf{Y} = (y_1, y_2, \ldots, y_t)$, the age at time t captures information from the past ages $(y_1, y_2, \ldots, y_{t-1})$.

In the LSTM model, the conditional probability $P(y_t|\mathbf{x}_t, y_1, y_2, \ldots, y_{t-1})$ is calculated by

$$P(y_t|\mathbf{x}_t) = \arg\max P(y_t|\mathbf{x}_t, y_1, y_2, \ldots, y_{t-1}). \tag{5}$$

The probability of sequence $\mathbf{Y}$ is calculated by multiplying conditional probabilities,

$$P(\mathbf{Y}) = \prod_{i=1}^{t} P(y_i|\mathbf{x}_i, y_1, y_2, \ldots, y_{i-1}), \tag{6}$$

where t is a person's the number of face images.

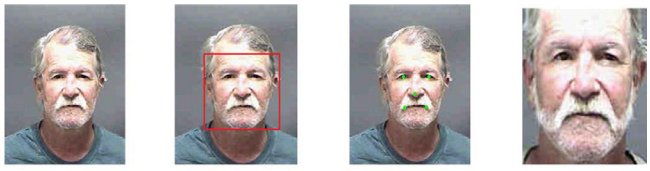
According to LDL, the output of sequence $\mathbf{Y}$ is ground-truth sequential label distributions $\mathbf{D} = (\mathbf{D}_{\mathbf{x}_1}, \mathbf{D}_{\mathbf{x}_2}, \ldots, \mathbf{D}_{\mathbf{x}_t})$. The probability of sequence $\mathbf{D}$ is calculated by multiplying conditional probabilities,

$$P(\mathbf{D}) = \prod_{i=1}^{t} P(\mathbf{D}_{\mathbf{x}_i}|\mathbf{x}_i, \mathbf{D}_{\mathbf{x}_1}, \mathbf{D}_{\mathbf{x}_2}, \ldots, \mathbf{D}_{\mathbf{x}_{i-1}}). \tag{7}$$

Let $\mathbf{D}^* = (\mathbf{D}^*_{\mathbf{x}_1}, \mathbf{D}^*_{\mathbf{x}_2}, \ldots, \mathbf{D}^*_{\mathbf{x}_t})$ represent the predicted sequence distribution. Our goal is to create a predictor $P : X \rightarrow \mathbf{D}^*$, where $\mathbf{D}^*$ is similar to $\mathbf{D}$. The total loss of cross entropy (8) is used to optimize the LSTM model,

$$Loss = -\frac{1}{t}\sum_{i=1}^{t} \mathbf{D}_{\mathbf{x}_i} \log \mathbf{D}^*_{\mathbf{x}_i}. \tag{8}$$

LSTMs networks learn the information of long-term memory spontaneously. A LSTM unit is equipped with three crucial gate structures, i.e., input gate, forget gate, and output gate, to protect and control information. Moreover, update state is used to update old state, and memory cells can remember information for long steps.



(a) Original image (b) Detection (c) Facial point (d) Alignment

Fig. 7. Overview of face image preprocessing, (a) original face image, (b) face is detected, (c) five facial landmarks are located, and (d) face is aligned with respect to five landmarks.

4. Experiments and results

4.1. Databases

We use two public databases MORPH [29] and FG-NET[1] to test the performance of our algorithm. MORPH contains 55,132 face images, of which more than 13,000 volunteers, with the maximum age of 77, and the minimum of 16. Each people has four images on average. Moreover, MORPH possesses underlying sequential patterns. FG-NET contains 82 persons. Each person has 12 images on average, the ages from 0 to 69.

4.2. Baselines

IIS-LLD and CPNN. Geng et al. [16] proposed the LDL with hand-crafted method for age estimation. Two different LDL algorithms i.e., IIS-LLD and CPNN algorithms are used to predict age.

IIS-ALDL and BFGS-ALDL. In order to improve IIS-LDL, Geng et al. [14] proposed two adaptive LDL, ie.,IIS-ALDL and BFGS-ALDL.

AGES. Geng et al. [17] proposed AGES with hand-crafted method. They established the aging pattern for each individual. The image's position in the subspace corresponding to the person's age.

DLDL. Gao et al. [12] combined CNN (VGG-16) with LDL to estimate age.

4.3. Image preprocessing

By image preprocessing, the background of an image is removed and the interference of irrelevant information is reduced. We employ the DPM model [27] to detect facial regions. As shown in Fig. 7, the process is divided into four stages, including preparing the original image, detecting the bounding box, checking point, and alignment. The cropped small-scale face image contains more texture information that is closely related to the human age. In order to feed into cascaded CNNs, all cropped face images are transformed to 224*224 pixels, and each image is represented by three channels.

4.4. Evaluation criteria

The performance of RAE is evaluate by two criteria, one is mean absolute error (MAE), and another is cumulative score (CS).

MAE is calculated by:

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |\, y - y^* \,|, \tag{9}$$

[1] http://www-prima.inrialpes.fr/FGnet/.

Table 1
Parameter setting on MORPH and FG-NET in CNN.

	MORPH	FG-NET
dropout rate	0.8	0.8
learning rate	0.0001	0.0001
epoches	20	100
mini-batches	80	2

Table 2
Parameter setting on MORPH and FG-NET in LSTM.

	MORPH	FG-NET
learning rate	0.0001	0.0001
epoches	200	200
mini-batches	80	2

Table 3
Comparison of MAE based on MORPH and FG-NET .

Method	MORPH	FG-NET
IIS-LLD	5.67	5.77
CPNN	4.87	4.76
BFGS-ALDL	4.34	
IIS-ALDL	4.43	
AGES	8.07	6.22
DLDL	2.43	3.76
RAE	1.32	2.19

where y indicates age which is labeled in database, y^* indicates the estimated age from the model.

CS_l is the accuracy rate of correct estimation, and the definition of CS_l is:

$$CS_l = \frac{N_l}{N} \times 100\%, \tag{10}$$

where N_l is test images which satisfying $|y - y^*| \leq l$, and l is the set years.

4.5. Training details and discussions

For the CNN and LSTM models, all weights are initialized to small random values sampled from Gaussian distribution, the mean and standard deviation are 0 and 0.001, respectively. All biases are initialized to 0.

For the CNN model structured based on Inception v4 (which tends to use a deeper architecture and smaller convolution filters), we train the Inception v4 using MORPH and FG-NET respectively. The values of all parameters in CNN are listed in Table 2. 1536-dimensional features from each image are extract by final dropout layer.

For the LSTM model, sequential features of each person are fed into LSTM. The label distribution is generated by Eq. (3). We find the best value $\sigma = 3.0$. MORPH and FG-NET are randomly divided into two parts according to the index of people, 80% for training and 20% for testing. The values of all parameters in LSTM are listed in Table 3.

The results of RAE and the baselines are compared in Table 1. We can see that RAE is significantly better. On MORPH, the MAE value of RAE is 1.32, compared to IIS-LLD of 5.67, CPNN of 4.87, BFGS-ALDL of 4.34, IIS-ALDL of 4.43, AGES of 8.07 and DLDL of 2.43. On FG-NET, the RAE value of MAE is 2.19, compared to IIS-LLD of 5.77, CPNN of 4.76, AGES of 6.22 and DLDL of 3.67. Figs. 8 and 9 show different CS curves on MORPH and FG-NET respectively. Fig. 8 reveals that the result of RAE on MORPH is the most accurate. The result on FG-NET demonstrated in Fig. 9 is similar to Fig. 8. Note that for DLDL, the accuracy rate is still significantly higher than IIS-LLD, CPNN, IIS-ALDL, BFGS-ALDL and AGES which extract features using hand-crafted method. We can intuitively conclude that deep learning methods are superior to traditional manual methods.

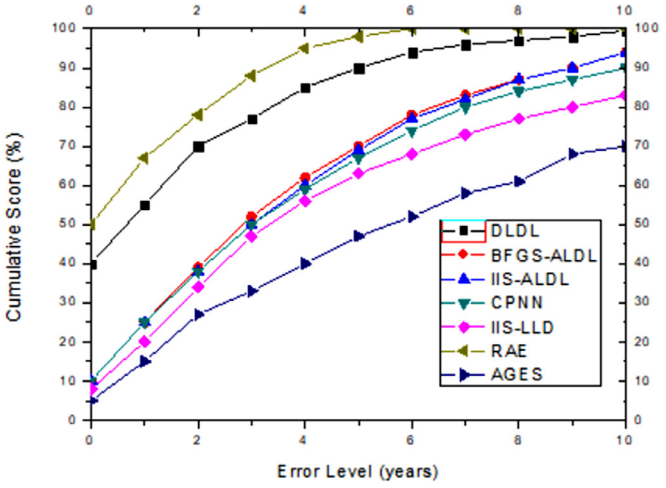


Fig. 8. Comparisons of CS based on MORPH.

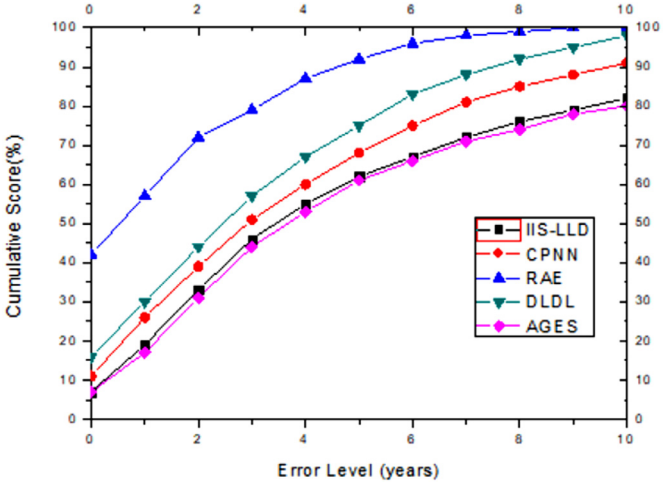


Fig. 9. Comparisons of CS based on FG-NET.

RAE obtains much better results than DLDL because RAE takes advantage of personalized aging patterns learned by LSTMs. In addition, the improvement in MORPH is obviously higher than that of FG-NET, and we think that the main reason is that MORPH is larger than FG-NET.

We investigated the computational time with RAE and DLDL on MORPH and FG-NET, the networks were built on a GPU with Nvidia Titan Xp. Table 4. shows computational time, feature dimension, testing time for one face and epoches of RAE and DLDL on MORPH and FG-NET. For RAE, the computational time includes training CNN model, image feature extraction and training LSTM model. From the results, our method RAE is slower than the DLDL, however, it is a good trade-off when both the age estimation accuracy and speed are concerned. Moreover, RAE can meet real-time requirements when a GPU is used to extract features.

Through the experimental results, we conclude the reasons for the excellent results of RAE. (1) The CNN of Inception v4 is a kind of excellent neural network, and its feature extraction has obvious discriminative property. Accurate features lay a solid foundation for learning aging patterns; (2) Personalized aging patterns enhance

Table 4
Feature dimension, computational time of RAE compared with DLDL .

Datasets	Methods	Feature dimension	Time	Testing time(ms)	Epochs
MORPH			16 h		
	RAE	1536		91	200
FG-NET			1.5 h		
MORPH			2.1 h		20
	DLDL	4096		7	
FG-NET			0.2 h		100

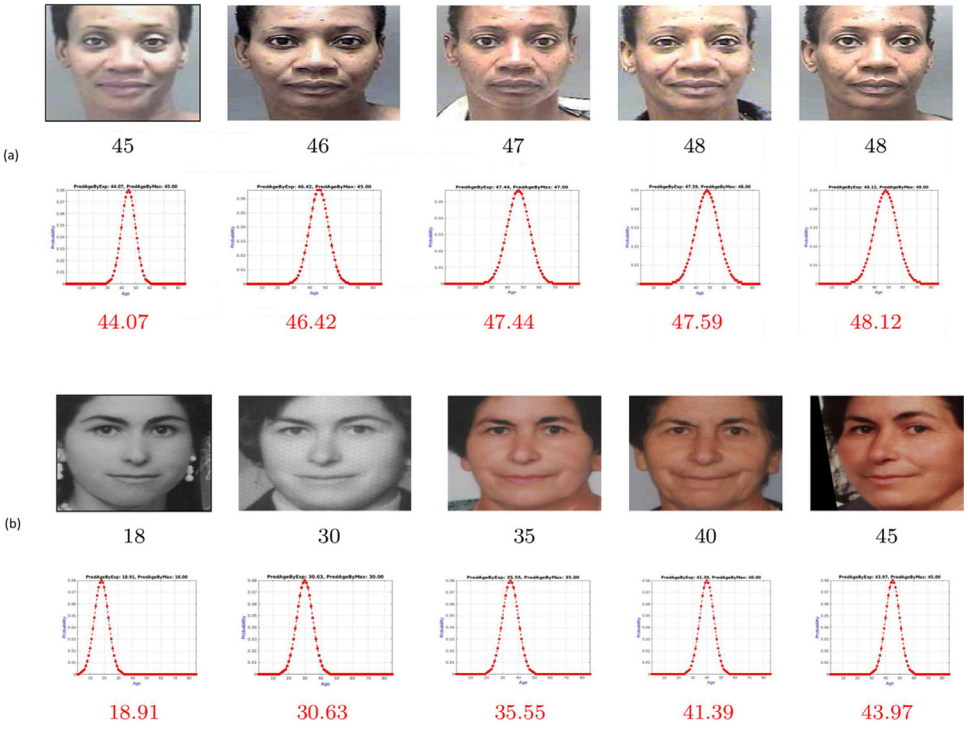


Fig. 10. Examples of two subjects' face images and their predicted ages. For (a) and (b), it shows five preprocessed faces and their real ages from MORPH and FG-NET, corresponding to the estimated age and distribution below them.

the performance of age estimation. Recurrent learning with the remember ability is the key to learning the aging patterns. LSTMs are able to effectively learn personalized aging patterns based on sequential features; (3) The extended LDL is used to overcome the obstacle of insufficient training set. The label distribution $\mathbf{D}_{\mathbf{x}_n}$ contains several none-zero element values, and the rich label distribution helps reduce over-fitting.

In Fig. 10, we provide several examples of face images with real ages from the two databases and predicted label distributions by RAE. The results of predicted ages are close to the real age. In addition, the entire prediction accuracy is gradually improved with the increase of sequential images. But when weak dependency images, such as old photos and facial postures, are added to the image sequence, the prediction accuracy is not significantly improved.

5. Conclusion

We propose RAE which combines CNNs and LSTMs for age estimation. In RAE, CNNs are used to extract facial image features, and sequential features are input into LSTMs to learn personalized aging patterns and generate ages.

In addition, we integrate LDL to LSTMs, LDL to exploit ambiguity from the real and adjacent ages. The label ambiguity can effectively overcome the problem of insufficient training images. Experimental results on MORPH and FG-NET have shown that RAE produces robust and competitive performances than current approaches.

In order to train LSTMs, we need databases with personal sequential images. Many small databases, however, do not have personal sequential images. Although the proposed RAE is effective, we believe that the advantage of RAE will further increase the accuracy when personal training set sizes grow.

Acknowledgments

The authors thank the financial support of the China National Natural Science Foundation (61702095), Anhui Polytechnic University Scientific Research Foundation (S031702004), Natural Science Foundation of Fujian Province (2018J01806) and Scientific Research Program of Outstanding Talents in Universities of Fujian, Natural Science Foundation(njpj2019209) of Nanjin Tech University Pujiang Institute.

References

[1] E. Arisoy, A. Sethy, B. Ramabhadran, Bidirectional recurrent neural network language models for automatic speech recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, Queensland, Australia, 2015, pp. 5421–5425.
[2] M. Aydogdu, V. Celik, M. Demirci, Comparison of three different cnn architectures for age classification, in: IEEE 11th International Conference on Semantic Computing,San Diego, California, USA, 2017, pp. 372–377.

[3] I. Bouchrika, A. Ladjailia, N. Harrati, Automated clustering and estimation of age groups from face images using the local binary pattern operator, in: The 4th International Conference on Electrical Engineering, Boumerdes, Algeria, 2015, pp. 1–4.
[4] S. Chen, C. Zhang, M. Dong, Deep age estimation: from classification to ranking, IEEE Trans. Multimedia 20 (8) (2017) 2209–2222.
[5] T. Cootes, G. Edwards, C. Taylor, Active appearance model, Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.
[6] M. Dehshibi, A. Bastanfard, A new algorithm for age recognition from facial images, Signal Process. 90 (8) (2010) 2431–2444.
[7] J. Donahue, L. Hendricks, M. Rohrbach, Long-term recurrent convolutional networks for visual recognition and description, Pattern Anal. Mach. Intell. 39 (4) (2017) 677–691.
[8] Y. Dong, Y. Liu, L. S.G., Automatic age estimation based on deep learning algorithm, Neuro Comput. 187 (2016) 4–10.
[9] M. Duan, K. Li, K. Lia, An ensemble cnn2elm for age estimation, IEEE Trans. Inf. Forensics Security 13 (3) (2017) 758–772.
[10] Y. Duan, J. Lu, J. Feng, Context-aware local binary feature learning for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (5) (2018) 1139–1153.
[11] Y. Fu, T. Huang, Human age estimation with regression on discriminative aging manifold, IEEE Trans Multimedia 10 (4) (2008) 578–584.
[12] B. Gao, C. Xing, Deep label distribution learning with label ambiguity, IEEE Trans. Image Process. 26 (6) (2017) 2825–2838.
[13] X. Geng, R. Ji, Label distribution learning, Trans. Knowl. Data Eng. 28 (7) (2014) 1734–1748.
[14] X. Geng, Q. Wang, Y. Xia, Facial age estimation by adaptive label distribution learning, in: The 22nd International Conference on Pattern Recognition,Stockholm, Sweden, 2014, pp. 4465–4470.
[15] X. Geng, Y. Xia, Head pose estimation based on multivariate label distribution, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1837–1842.
[16] X. Geng, C. Yin, Z. Zhou, Facial age estimation by learning from label distributions, IEEE Trans. Pattern Anal. Mach. Intell. 35 (10) (2013) 2401–2412.
[17] X. Geng, H. ZhouZ, K. Smithmiles, Automatic age estimation based on facial aging patterns, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2234–2240.
[18] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press., 2016.
[19] G. Guo, Y. Fu, C. Dyer, Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 17(7) (2008) 1178–1188.
[20] G. Guo, G. Mu, Y. Fu, Human age estimation using bio-inspired features, in: IEEE Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 112–119.
[21] Z. Hu, Y. Wen, J. Wang, Facial age estimation with age difference, Trans. Image Process. 26 (7) (2017) 3087–3097.
[22] Y. Kwon, N. Vitoria Lobo, Age classification from facial images, in: IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 1999, pp. 762–767.
[23] H. Liu, J. Lu, J. Feng, Ordinal deep learning for facial age, Estimation IEEE Trans. Circuits Syst. Video Technol. 29 (2) (2019) 486–501.
[24] H. Liu, J. Lu, J. Feng, Label-sensitive deep metric learning for facial age estimation, IEEE Trans. Inf. Forensics Security 13 (2) (2018) 292–305.
[25] J. Lu, V. Liong, J. Zhou, Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (8) (2018) 1979–1993.
[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] M. Mathias, R. Benenson, M. Pedersoli, Face detection without bells and whistles, in: European Conference on Computer Vision, Zurich, Switzerland, 2014, pp. 720–735.
[28] C. Ng, M. Yap, Y. Cheng, Hybrid aging patterns for face age estimation, Image Vis. Comput. 69 (2017) 92–102.
[29] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 2006, pp. 341–345.
[30] J. Wang, G. Su, X. Lin, Age estimation from facial images, J. Tsinghua Univ. 47 (4) (2007) 526–529.
[31] J. Wang, Y. Yang, J. Mao, Cnn-rnn: a unified framework for multi-label image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, United States, 2016, pp. 2285–2294.
[32] X. Wang, R. Li, Y. Zhou, A study of convolutional sparse feature learning for human age estimate, in: The 12th IEEE International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 2017, pp. 566–572.
[33] Z. Zhang, M. Wang, X. Geng, Crowd counting in public video surveillance by label distribution learning, Neurocomputing 166 (1) (2015) 151–163.