Neurocomputing 187 (2016) 4–10

Contents lists available at ScienceDirect

Neurocomputing

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

Automatic age estimation based on deep learning algorithm

Yuan Dong [a], Yinan Liu [a], Shiguo Lian [b,*]

[a] Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[b] Huawei Central Research Institute, Beijing 100086, PR China

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form
13 September 2015
Accepted 14 September 2015
Available online 30 November 2015

Keywords:
Age estimation
Deep Convolutional Neural Networks
Transfer learning
Loss function

ABSTRACT

Automatic age estimation has attracted much attention due to its potential applications. Most of the proposed approaches have mainly used low-level handcraft features to encode facial age related visual information and train an age estimation model. In this paper, we focus on age classification task in which face image is assigned to a label that represents an age range. We proposed a deep learning based framework for age classification task. In our proposed algorithm, Deep Convolutional Neural Networks (Deep ConvNets) are used to extract high-level complex age related visual features and predict age range of input face image. Due to lack of age labeled face images, we use the transfer learning strategy to train the Deep ConvNets. In addition, to describe the relationships between labels that compose an ordered sequence, we define a new loss function in the training process of age classification task. The experiments are conducted on a widely used age estimation dataset-Images of Groups of People. The experimental results demonstrate the excellent performance of our proposed algorithm against the state-of-the-art methods.

1. Introduction

Automatic age estimation task aims to use machine learning algorithms to estimate a person's age based on features extracted from face image. As an important attribute of face, age estimation has attracted much attention due to its potential applications in Human–Computer Interaction (HCI) [1], electronic customer relationship management [2], surveillance monitoring [3] and so on. As a branch of face recognition, automatic age estimation task also shares several similar processes such as face detection, face alignment, feature extraction and estimation with other facial related tasks. Features extracted from face images representing age related visual information is a key factor to the performance of age estimation algorithms. Thus, many age-related feature extraction algorithms have been researched such as anthropometric models [4], active appearance model (AAM) [5], age pattern subspace and age manifold [6]. Low-level handcraft features are usually extracted in these typical algorithms. Given features, automatic age estimation tasks can usually be categorized to two classes: age classification task [6,7] and age regression task [8–10]. In age classification task, face image is usually assigned to a class label that represents an age range and multi-class classification algorithms are always used. In age regression tasks, algorithms output an accurate predicted age number. In this paper, we focus on the classification task that assigns a face image to an age range label.

Deep learning algorithms develop rapidly recently and have achieved success in many pattern recognition tasks. In ILSVRC 2012, Krizhevsky et al. proposed a Deep ConvNets structure for 1000 classification task and get the top performance [12]. Simonyan and Zisserman proposed a deeper and wider structure compared with Alex-Net to improve the classification performance [13]. Szegedy et al. also increased the depth and width of CNN for classification and detection task [14]. Some researchers have applied deep learning algorithms to other pattern recognition tasks such as speech recognition [15–17] and pedestrian detection [18–20]. These experiments demonstrate that high-level semantic features extracted by deep learning algorithms usually perform better than handcraft features. However, as introduced in the first paragraph, traditional age estimation methods usually use low-level handcraft features to train a model. So, in this paper, we propose a novel age estimation method based on Deep ConvNets. We train Deep ConvNets to get high-level facial features to improve the age estimation performance.

The Deep ConvNets in our proposed algorithm is trained as a multi-class classifier. The public available age estimation datasets usually do not contain enough labeled images to train Deep ConvNets that owns hundreds of millions of parameters. We take the

* Corresponding author.
E-mail addresses: yuandong@bupt.edu.cn (Y. Dong), liuyinan1988@126.com (Y. Liu), shiguo.lian@ieee.org, sg_lian@163.com (S. Lian).

http://dx.doi.org/10.1016/j.neucom.2015.09.115
0925-2312/

transfer learning strategy to overcome this problem. Transfer learning has been widely used in some image classification tasks and semi-supervised learning tasks [21–23]. It has been proved to be very efficient to train models on small scale dataset. We first train Deep ConvNets for face identification task with a large scale dataset. And the parameters in trained models are used as initialization for age classification task. In age classification task, age labels are ordered. For example we define 3 labels: label_1, label_2 and label_3, that represent age range 0–2, 3–7 and 8–12, respectively. The distance between label_3 and label_1 should be shorter than that between label_2 and label_1. In other words, label_2 is more related to label_1 than label_3. However normal multi-class classification algorithms such as the Soft-max do not take relationships between labels into account. To improve performance, in our proposed algorithm, we define a new loss function in which a distance term is added to describe the relationships of labels. In addition, in many face related tasks, combination of results produced by different regions in face image has been proved to perform much better than single region or the whole face image. Thus, we train multiple independent Deep ConvNets models with patches cropped from face images and combine the results as the final results to get a better performance. We conduct the experiments on widely used age estimation dataset-Images of Groups [9]. The proposed algorithm demonstrates excellent performance compared to the state-of-arts.

2. Related work

Recently, some researchers have applied deep learning algorithms to face verification tasks or other face related tasks. Sun et al. proposed a DeepID structure to extract discriminative face features for face verification task [24]. The DeepID algorithm is improved by adding a verification constraint in loss function to get better performance in [25]. A cascaded Deep ConvNets structure is proposed to detect face landmark points in [26]. A deep multi-task learning algorithm is proposed to detect landmark points in [27]. In this paper, we proposed a framework based on Deep ConvNets to automatically estimate age. We take similar structure with DeepID [24] that extracts discriminative face features for face verification tasks.

In DeepID, 60 patches are cropped from face images and each patch is used as input data to one independent network. Thus, 60 independent networks are trained by different patches from face images. Structure of networks in DeepID is shown as Fig. 1. Input size of network is 39×31 pixels. Each network contains 4 convolutional layers, one fully-connected layer and one output layer. Each of the first three convolutional layers is followed by one Max-pooling layer. Max-pooling layer is formulated as

$$y_c(i,j) = \max_{0 \le m,n \le s} \{x_c(i \bullet s + m, j \bullet s + n)\} \tag{1}$$

where x is input feature maps of Max-pooling layer and c is the index of feature maps and s is the window size of Max-pooling operation. The last convolutional layer is locally connected that means convolutional weights and biases in this layer are not shared in different positions. The fully-connected layer is connected with the last Max-pooling layer and the last convolutional layer. The last layer is a Soft-max layer. Soft-max layer produces a probability for each class label. In DeepID, the networks are trained to classify face images from 10,000 persons and each person is assigned to a class label. The fully-connected layer's neuron activations are extracted as feature vector of input image. There are 60 networks in DeepID, so each input image produces a 9600-dimensional feature vector, and the feature vector is used to train a classifier for face verification task. We take similar structure with DeepID for age classification task. We do not extract features and train classifier independently but we use the output of CNNs as age estimation result.

3. Architecture of our proposed algorithm

3.1. Transfer learning

In this paper, we train Deep ConvNets to predict age range of input image. We first detect the bounding box of face and detect the five key points (two eyes, nose and two mouth corners) with deep cascade networks algorithm in [26]. Face image is aligned by similarity transformation with key points. Current available age estimation datasets usually do not contain enough labeled images to train the Deep ConvNets. To overcome this problem, we take transfer learning strategy in our proposed algorithm. As shown in Fig. 2, transfer learning includes two processes: pre-train and fine-tune. In pre-train process, the randomly initialized networks are first trained by a related task that owns enough labeled images. In fine-tune process, parameters learnt in pre-train process are used as initialization for new task. In our proposed algorithm, we first train networks for a face identification task that owns enough labeled face images. Networks trained by face identification task are able to extract discriminative facial features. Then we fine tune parameters in the trained networks for age estimation task.

3.2. Multiple models

In our proposed algorithm, we crop 2 centered regions from original face image on 2 scales and each scale produces 10 patches (the whole image, 4 horizontal strips and 5 regions surrounding 5 key points). In Fig. 3, the first row shows original face image, face

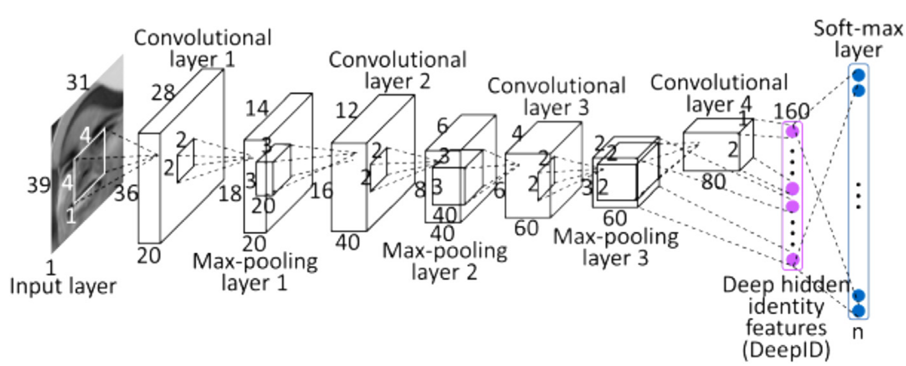

Fig. 1. Network structure of DeepID.