ATLANTIS PRESS
Advances in Computer Science Research, volume 71
4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA 2016)

Apparent Age Estimation with CNN

Zhiqin Zhang[1, a]
[1]School of computer science, Wuhan Donghu University, Wuhan 430000, China;
[a]120991228@qq.com

Keywords: age estimation, convolutional neural networks, IMDB-WIKI dataset .

Abstract. Apparent age estimation from face image has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. In this paper ,we tackle the estimation of apparent age in still face images with deep convolutional neural networks (CNN).Our convolutional neural network use the GoogLeNet architecture, add batch normalization layer after each ReLU operation and remove all the dropout operations to accelerate the convergence of this very large-scale deep network. In addition, due to the limited number of apparent age annotated images, we train the deep models with several datasets in a cascaded way. Firstly, We pre-train a real age estimation model using IMDB-WIKI dataset, and then fine-tune the deep model with combined dataset with multiple real-age labeled databases .Finally, the apparent age data from the challenge are used to fine-tune the deep model parameters for apparent age estimation.

1.Introduction

Age estimation has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media.And it has drawn increasing attention in computer vision with its potential applications on video surveillance, access control, and demography [1, 2].

There are numerous studies and several large datasets on the (biological, real) age estimation based on a single face image.Gil Levi[3] applied convolutional neural networks (CNN) to age estimation for the first time. Wang [4] used a 7-layers deep convolutional network to learn deep age patterns, followed by both SVR and CCA for final age estimation. Different from those traditional methods that use local features for representing face images,the above methods made tremendous progress by the use of deep convolutional neural networks (CNN).

In practice, the very large-scale deep convolutional neural network has yielded quite impressive performance in image recognition problem [5].For good performance, more and more scientists use the large CNN architectures with large datasets[6,7,8,9].[6] used a multi-scale deep convolutional neural network for fully end-to-end age regressor learning. Rasmus Rothe[7] used the VGG-16 architecture [10] . Yu Zhu et.al [7] made their network based on GoogLeNet[5]. Xin Liu[9] combined classification models and regression models, both exploiting very large-scale deep convolutional neural network.These methods have achieved better performance benefined from the deep learning technology.

Recently, apparent age estimation from face image has attracted more and more attentions as it is favorable in some real world applications,and becomes a new measurement towards real age.As defined, apparent age focus on predicting“how old does the person look like?”. Thus, apparent age is labeled by different volunteers given only the images containing the single individuals,Such as the apparent age data is from the ChaLearn Looking At People (LAP)[11]. A key difference between apparent age estimation and the traditional age estimation is that the age labels are annotated by human assessors rather than the real chronological age. In reality, some people may look younger than the real chronological age, while some may look older. As a result, the apparent age may be similar but different from the real age. Rasmus Rothe[7] made great contribution for apparent age estimation by the use of finetuning on LAP dataset.

The problem of apparent age estimation is studied in this work. Particularly, we firstly utilize a large number of face images with real age labels to learn a real age representation using the deep

neural networks, then we study how to fine-tune the deep networks using a limited number of data with apparent age labels.Figure 1 demonstrates the general idea of the proposed method.

In our study, the apparent age data is from the ChaLearn Looking At People (LAP) challenge 2016. Using the samples by the challenge competition, we can achieve an error of 0.294835 on the final evaluation, and our result is ranked the 3rd place in this competition.

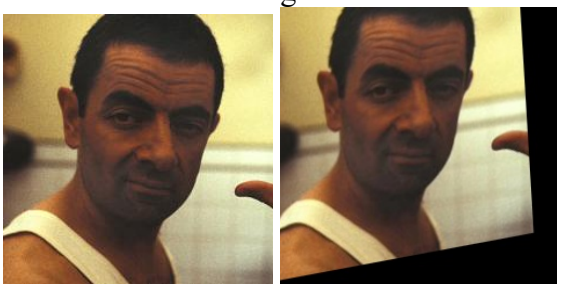
The rest of paper is organized as follows. Section 2 describes the details of our proposed method for apparent age estimation. Section 3 presents the databases that are used for training the deep networks. And Section 4 shows the experimental evaluations of the proposed method.

2.Proposed Method

2.1Face Preprocessing.

The face images for age estimation are preprocessed with two steps including face detection and face alignment. In face detection stage, we adopt the face detection toolkit developed by VIPL lab of CAS [12]. In training phase, Images are rotated every 10 degrees for further detection if no face is detected in the original image. The rest of images that are still not detected by rotation are not used in our approach.

Facial landmarks are important for good face recognition algorithms, especially for unconstrained face recognition problem. Based on the precise facial landmarks, one can correct the pose of face image or build pose robust face descriptors. In face alignment stage, We apply the Coarse-to-Fine Auto-Encoder Networks (CFAN) [13] to detect the five facial landmarks in the face the left and right center of the eyes, the nose tip, the left and right corner of mouth, and then perform alignment using method in [14].The alignment results are shown in figure 1.



(a)face image (b)aligned face image

(c)face image (d)aligned face image

Figure 1. Results of preprocess

2.2Deep Convolutional Neural Network.

Recently, the very large-scale deep convolutional neural network has yielded quite impressive performance in image recognition problem [5] in practice. The convolutional neural networks (CNN) have shown promising performance in age estimation. For good performance, we use the large CNN architectures with large datasets.

In our proposed method, we chose the GoogLeNet [11], which is 22-layer deep neural network, to train our deep models. We make two modifications of the GoogLeNet. First, we remove the two auxiliary loss layers. Second, we add batch normalization layer before each ReLU operation and remove all the dropout operations to accelerate the convergence of this very large-scale deep network. Owe to the usage of batch normalization, we found it unnecessary to add two auxiliary loss layers for the purpose of avoiding vanish of gradient problem. By removing the two auxiliary loss layers, the performance is even improved slightly.

For good performance, usually the large CNN architectures need large training datasets. However, the publicly available face image datasets are often of small to medium size, and often with real age not apparent age information .We make a good use of fine-tune. Figure 2 shows the three stages of this scheme,i.e., pre-train with large dataset of real age, fine-tune with small dataset of real age,and fine-tune with apparent age.

The first step is to pre-train the network using a large number of face images, e.g., from the IMDB-WIKI dataset [7]. In order to learn more robust and representative features by deep neural networks, in our method we expect to utilize GoogLeNet , [11] which is 22-layer deep neural network, to train our deep network. We hope that through this pre-training step, the deep network is able to capture general facial representations.

The second step for training our deep model is finetuning the network parameters using a large number of data with biological age labels. We merged multiple age databases into one training dataset, details of the databases we used is presented in Section 3.

Finally, we employ apparent age training set to fine-tune the deep network from step 2, producing a robust apparent age deep network. Literature References

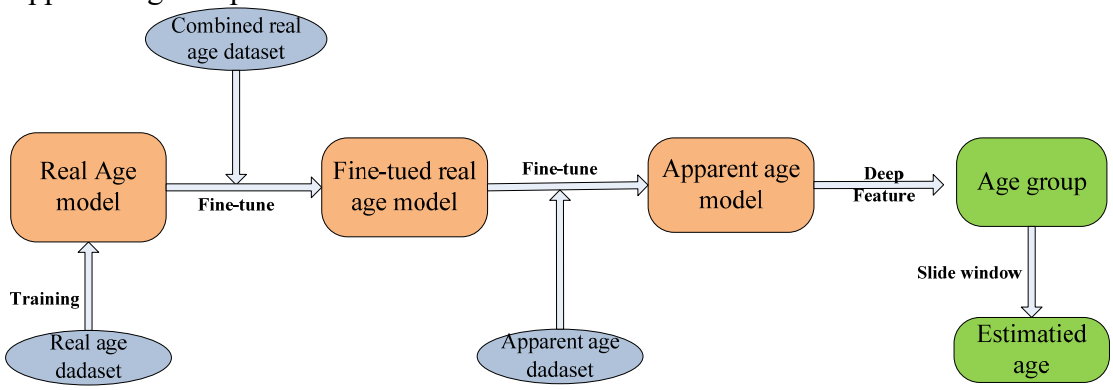


Figure 2 Illustration of out method for training deep model for apparent age estimation

2.3Age Classifier and Age Estimation.

Age estimation can be seen as a piece-wise regression or, alternatively, as a discrete classification with multiple discrete value labels. In our work, it is a one dimensional regression problem with the age being sampled from a continuous signal ([0,100]).

For improving the accuracy of the prediction,as shown in Fig. 3, we compute a softmax expected value based on glide window. In our case,after given the softmax probability of the 101 classes dimensional output layer,we slide a window on the 101 to find the group with max probability.Finally,we use the center of window as the expected value of age estimation.

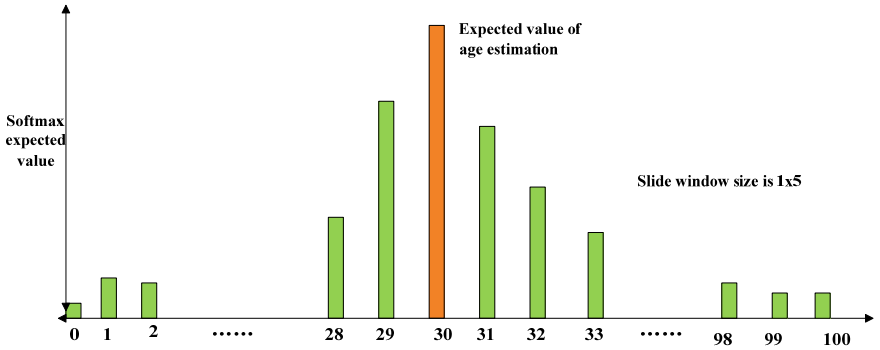


Figure 3 age estimation based on slide window

3.Datasets

3.1Adience Age Dataset.

Adience Age Dataset [3] is a small dataset including about 18,300 images. The sources of the images included in Adience Age Database are Flickr albums, assembled by automatic upload from iPhone5 (or later) smart-phone devices, and released by their authors to the general public under the

Creative Commons (CC) license. It's intended to be as true as possible to the challenges of real-world imaging conditions. In particular, it attempts to capture all the variations in appearance, noise, pose, lighting and more, that can be expected of images taken without careful preparation or posing.

However, this database doesn't have the accurate age annotation, i.e, age is annotated in a range form like 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60+. As we use Euclidean loss for the deep networks, in this work we transfer the range annotation into traditional single label by just taking the mean value, e.g., 28-32 is converted to 30.

3.2IMDB-WIKI dataset.

IMDB-WIKI dataset[8] is a large scale database which connected from IMDB website and Wikipedia. It took the list of the most popular 100,000 actors as listed on the IMDB website, and (automatically) crawled from their profiles birth dates, images, and annotations. By assuming that the images with single faces are likely to show the actor and that the time stamp and birth date are correct, the biological (real) age of each such image can be assigned. In total ,there are 461,871 face images from IMDB. From Wikipedia, it crawled all profile images from pages of people and filtered them according to the same criteria applied for the IMDB images. In total ,there are 62,359 images.

However, the accuracy of the assigned age information was not vouched.Besides wrong time stamps, many images are stills from movies,movies that can have extended production times. in addition,as some of the images (especially from IMDB) contain several people,the assigned mistakes were happened. For good performance in training phase, we cleaned the dataset by removing the apprently wrong assigned imges. This leaves us with 240,703 training images for our CNNs.

In this work, we use them as the training data of the first step, network pre-training, i.e., identity discriminant learning.

3.3LAP dataset.

The ChaLearn LAP dataset [11] consists of 4699 face images whice have the apparent age labels.As we know,A key difference between apparent age labels and the traditional age labels is that the apparent age labels are annotated by human assessors rather than the real chronological age.To reduce the error,each label in the dataset is the averaged opinion of at least 10 independent users. Therefore, a standard deviation σ is also provided for each age label.

However,The LAP has uneven age distribution ,which covers the 20-40 years interval best, while for the [0,15] and [65,100] intervals it suffers from small number of samples per year.In order to have even distribution samples for training ,we download some images of children and old people.

4.Experiments

4.1ICCV2016 Apparent Age Estimation Challenge.

The ICCV2016 Apparent Age Estimation Challenge aims to investigate the performance of estimation methods on apparent age rather than real age. A dataset of 4699 images is provided, each image is labeled a real number from 0 to 100 indicating the apparent age. The images are collected from two web-bases application and labeled by at least 10 different users.The ground truth is calculated with all the users' votes, where the mean age from all the votes is considered as the apparent age label.

The challenge is composed of development phase and final test phase. In the development phase, the model is trained on the training set and evaluation is conducted on the validation set. In the final test phase, both the training set and validation set can be deployed for model training. The performance is measured by mean normalized error calculated as:

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N} 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

where ε is the error, N is the number of test samples, x is the predicted age, μ is the mean age (labeled apparent age) and σ is the standard deviation.

4.2Implementation Details.

The training of out network is implemented using the Caffe framework. we set the base learning rate as 0.01 and reduce the learning rate by polynomial with gamma value equals to 0.5. The momentum is set as 0.9 and the weight decay is set as 0.0005. All the experiments was performed on an GTX-980 GPU machine with 8GB memory.

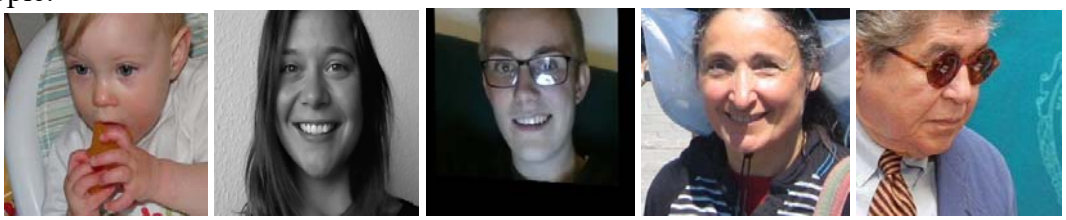
In stage of face preprocessing,each face was aligned to 256x256.

In stage of training,we pretrain our network using IMDB-WIKI dataset, which contains about 260,000 face images. And then, we combine other datasets (Adience,FGNET ,Morph-II) to a large dataset with real age labels.Fine-tune is performed on the combined dataset.In the last phase of training,we fine-tune our network on LAP to get an apparent age estimation model.

For the test phase,after given the softmax probability of the 101 classes,we find the group with max probability based on glide window which size was 1x5.

4.3Results.

Figure 4 presents some good and bad cases of apparent age estimation results by the proposed approach. It can be seen that our approach is robust to variations in pose, lighting, ethnicity, occlusion and color mode. However, our approach does not work very well for face blur, mis-alignment or senior people.



labels:1/est:1 labels:20/est:20 labels:20/est:20 labels:46/est:46 labels:62/est:62

(a) good cases of the proposed method

labels:1/est:40 labels:21/est:28 labels:21/est:29 labels:70/est:46 labels:72/est:80

(b) Bad cases of the proposed method

Figure 4 examples of age estimation results of the proposed method

5.Summary

In this paper, we propose a deep learning approach for robust apparent age estimation. For good performance,we deploy a very large-scale 22-layers deep convolution neural network with large datasets.To reduce the risk of over-fitting,we firstlytrained deep network with part real age images ,and then finetuning is presented on combined dataset with multiple real-age labeled databases ..Finally, the apparent age data from the challenge are used to fine-tune the deep model parameters for apparent age estimation.we have performed apparent age estimation in a coarse to fine manner. Finally, in the ChaLearn LAP challenge 2015 age estimation track, our method has shown promising results and achieved the 3rd place. In the future, we will explore more deep representations to improve the performance further.

Acknowledgment

This paper is supported by the Foundation for Young Scholars in Wuhan Donghu University in 2016(2016dhzk05).

References

[1] Y. Fu, G. Guo, and T. Huang. Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 32(11):1955–1976,2010.

[2] H. Han, C. Otto, X. Liu, and A. Jain. Demographic estimation from face images: Human vs. machine performance.IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 37(6):1148–1161, 2015

[3] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[4] X. Wang, R. Guo, and C. Kambhamettu. Deeply-learned feature for age estimation. In IEEE Winter Conference on Applications of Computer Vision (WACV), pages 534–541,2015

[5 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

[6] D. Yi, Z. Lei, and S. Li. Age estimation by multi-scale convolutional network. In Asian Conference on ComputerVision(ACCV), pages 144–158. 2015.

[7] Rasmus Rothe, Radu Timofte, Luc Van Gool. DEX: Deep EXpectation of apparent age from a single image , In IEEE International Conference on Computer Vision Workshop 2015:252-257

[8] Yu Zhu , Yan Li, Guowang Mu , and Guodong Guo 1 A Study on Apparent Age Estimation, In IEEE International Conference on Computer Vision Workshop, 2015:267-273

[9]Xin Liu, Shaoxin Li, Meina Kan.AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation, In IEEE International Conference on Computer Vision Workshop, 2015:258-266

[10]O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh,S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein,et al. Imagenet large scale visual recognition challenge.IJCV, pages 1-42, 2014. 1, 2, 4

[11] S. Escalera, J. Fabian, P. Pardo, X. Baro, J. Gonzalez, H. J.Escalante, and I. Guyon. Chalearn 2016 apparent age and cultural event recognition: datasets and results. In ICCV,ChaLearn Looking at People workshop, 2016. 1, 3, 4, 5, 6

[12] www.vipl.ict.ac.cn. Visual information processing and learning (vipl) group in institute of computing technology (ict) chinese academy of sciences (cas)

[13] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In European Conference on Computer Vision (ECCV), pages 1–16, 2014.

[14]Gary B. Huang, Marwan Mattar, Honglak Lee, Erik Learned-Miller.Learning to Align from Scratch Advances in Neural Information Processing Systems (NIPS), 2012