Survey of Techniques to Estimate the Age and Gender of A Person using Face Images

Naveen K B
Department of Computer Science and Engineering
B N M Institute of Technology
Bangalore, India

Dr. Vimuktha Evangeleen Salis
Associate Professor,
Department of Computer Science and Engineering
B N M Institute of Technology
Bangalore, India

Abstract— **Over the recent years, a great deal of effort has been made on age estimation and gender recognition from face images. An analysis of different methods to estimate the age and gender and transition of methodology over a period of time shows that approaches first taken involved Image Processing techniques such as wrinkle extraction using hough transforms. With the advent of Machine Learning, techniques such as linear SVMs were used to solve the problem of age/gender estimation. As computing power and the amount of available training data increased, deep neural networks such as Convolutional networks were developed to solve the problem with higher accuracy. The application of age and gender estimation in real- time and mobile devices paved way for the requirement of light- weight networks, which have low inferencing time on devices with lesser memory and computational power. MobileNets were modified to solve this problem, and ensured low inferencing time while at the same time maintaining high accuracy.**

I. INTRODUCTION

Age and gender play an important role in assessing a profile of a person. This area of research has developed due to applications such as criminal profiling, face recognition as a security measure, and better integration of artificial assistants in personal life(more recent). This survey aims at analyzing

- Techniques for age and gender estimation for low power embedded and general purpose computers
- The shift in approaches to solve this problem from non portable high compute systems with higher accuracy to portable embedded systems with acceptable accuracy

Applications such as criminal profiling and face recognition security systems require high accuracy results and can afford to use high compute systems to obtain the same. On the contrary artificial assistants or any application that requires age/gender estimation and has to run on portable devices will require power efficient solutions but can sacrifice on accuracy.

II. BENCHMARK DATASETS

The method used, and the results of an approach to classify the gender and age based on face images depend greatly on the dataset used to train the data. Listed below are some of the most commonly used benchmarks for the problem mentioned.

A. FG-Net Aging

The dataset has images of 82 subjects of about 1000 images each with accurate age for each subject. These photos were captured under controlled conditions therefore do not reflect real time scenarios and cannot be used for face recognition systems.[1]

B. Gallagher

This dataset consists of people posing for the camera. Multiple subjects are present in a forward facing manner. The median face included in the set occupied only about 18.5 pixels between centres of the eye and 25% of the faces occupy less than 12.5 pixels. The age estimation can be done because of the age labels corresponding to the images.[1]

C. Morph

It was a dataset collected by the Face Aging Group at University of North Carolina at Wilmington. The dataset has multiple subsets out of which Album 2 is freely available. This has photographs of about 13,000 individuals summing upto around 55,000 images.[1]

D. VADANA

This dataset is relatively recent and has images of 43 subjects and about 2,298 images in total. This provides multiple images of the same subjects across ages and hence helps us study age progression on a face image.

E. Adience

The images included in this collection attempt to capture all variations in noise, appearance, lighting, pose and more, that are a result of taking images without careful preparation and in an unconstrained manner. The dataset is intended to be as true as possible to the challenges of real-world imaging conditions, and contains 26,580 images of 2,284 subjects. Each image in the dataset has a gender label and belongs to one of 8 age groups.[1]

F. IMDb-wiki

It is currently the largest public dataset of face images with age and gender labels. It consists of 523,051 images of the most popular 100,000 actors as listed on the IMDb website. The images are crawled from their profiles, name, date of birth, gender and all images related to that person. [9]

III. METHODS USED

A. Age Estimation

Age estimation with face images involves extraction of features from a face. Anthropometry methods were used earlier to estimate the age[7] based on extracted features and see the proportions of features progress over time, like

the length of the forehead, et cetra. Features like wrinkles were used earlier where it was assumed that the concentration of wrinkles, and their depth on faces increased over time [1]. The prediction was made as an approximate age group that the person belonged to but the ideal approach is to predict a fixed value or to allow a much stricter error range.[1]proposes the first step to be the alignment of faces. Viola Jones face detector is used to detect the key points of a face and images are aligned based on a single reference coordinate. Facial feature detectors like Zhu and Ramanan[9] detects 68 specific features and based on that affine transformations like warping can be applied to the image. To overcome the errors we repeat the process of feature detectors and affine transformation with mean squared errors to get the least error prone match. Practically just one iteration gave good improvement in the results. Aligned faces are encoded using image representations like LBP. They use local binary patterns by modelling the high dimensional data onto a feature space to characterize changes of facial features over age. Gabor filter's are further used to analyze textures on the extracted features. Following this a dropout SVM was used for training purposes. A dropout SVM was inspired by the drop out principle in neural networks. For the multi-label age classification a one vs one linear-SVM is used. This is done to avoid over fitting.

In [8] human estimators were asked to estimate the gender and

age period(5 year range) of faces with hair and clothes, and edited images without hair and clothes. Though the latter had a lesser accuracy of 0.928 for gender and 0.880 for age as compared to the former with an accuracy of 0.990 for gender and 0.906 for age, both showed very high correlation between face image and age/gender. In [8] the skin region of the face is first extracted by using a colour image, and HSV values. Histogram equalization is used in to enhance the wrinkles on the face. The wrinkles are then extracted using noise reduction, edge detection, thinning and finally a special Hough transform, DTHT(Digital Template Hough Transform) [8]. The number of wrinkles obtained after extracting them through the above mentioned procedure is the basis for age classification. These values are referred in a lookup table made using the HOIP-FACE-DB. The proposed method achieved accuracy of around 0.27 for age.

B. Gender Estimation

Gender estimation using face images is done mostly with an additional module along with age. Gender estimation involves prediction of the likelihood of a face belonging to male or female. Reference [8] uses features such as the smoothness of the face contour, and the ratio between the sizes of face contour and of facial parts. This ratio is quite different among male and female. The PICASSO system is used to investigate the caricatures of men's faces and female faces. PICASSO calculates the average face of men, and exaggerates the features to form a caricature. The result of the investigation was that the facial contour of female is smoother than that of male. The fact that 20-30 year old female faces are covered in make-up affected the age estimation, and hence gender is first estimated,

followed by estimating the age [8].The proposed method achieved accuracy of around 0.87 for gender estimation. The approach followed by other references involve methods like SVM [1] where feature vectors of the face images are extracted and a linear SVM can classify whether the person in the image is male or female. [2].

Fig. 1. Images with difficulty to classify gender[2]

Fig. 2.   Images where age was miss-classified[2]

C. Simultaneous Age and Gender Estimation

Many approaches use a single CNN to estimate both age and gender, rather than using multiple CNNs each to predict the age and gender separately. This is done using a single output layer with multiple classes, 2 for the gender and the rest for the age classification[2], or a single model with 2 output layers, one for age classification and the other for gender classification. The introduction of Convolution Neural Networks into this domain improved the results[5]. CNN's can be used to estimate age as a number rather than an age group [2]. Due to the lack of a large data-set with age as a number this paper predicts age-groups instead of a specific age. The architecture proposed in [2] uses 3 convolution blocks each with different number of filters of different sizes. The model had a total of 10 output classes out of which 2 stood for gender and each value depicted the probability of being a male vs female. A CNN is compute intensive and requires high performance computers to give results in a real- time scenario. This is not viable for many of the applications that require age and gender estimation. Mobile Nets seem like an ideal way to go where the neural networks are designed to be highly power efficient with respectable results [4]. [3] uses MobileNets for the estimation of age and gender together. A MobileNet based architecture is referred to as a light weight CNN. MobileNets have features such as low

memory usage, hard parameter sharing and they have a 1x1 kernel size [6]. This model was built using the adience data set as a benchmark. [9] automatically extracts persons and their attributes (gender, year of born) from an album of photos and videos. A two-staged approach is used in which firstly the convolutional neural network simultaneously pre- dicts age/gender from all photos and additionally extracts facial representations suitable for face identification[9].The MobileNet, which is preliminarily trained to perform face recognition is modified in order to additionally recognize age and gender. In the second stage of our approach, extracted faces are grouped using hierarchical agglomerative clustering techniques. The born year and gender of a person in each cluster are estimated using aggregation of predictions for individual photos. The gender ecognition task is considered a binary classification problem, and the output layer is made a sigmoid. The age prediction is considered a special case of a regression problem, and is converted to a multi-class classificaton with N(the number of classes) = 100 so the predicted ages can be 1,2,..100 years old.A softmax output layer is used for this task. The data set used is a combination of Adience and IMDb-Wiki.

Age estimation sometimes can be very difficult and some-times its hard for us human to predict accurately. Some images from the adience dataset that are often predicted wrong are given as part of [Fig. 2].Mobile net architecture when incorporated with this specific task was tuned in such a way that it could simultaneously predict both age and gender as part of its output. There are few cases where gender estimation can be difficult even from a humans perspective and these are cases that are usually miss-classified [Fig. 1].
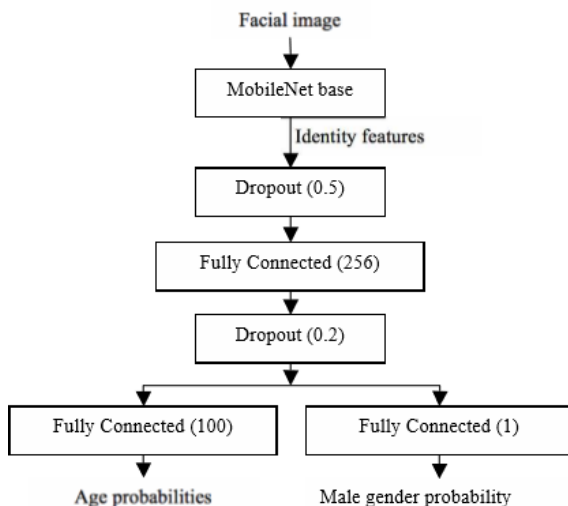

Fig. 3. Architecture of the modified MobileNet in [9]

IV. COMPARISON OF APPROACHES

We compare the different approaches discussed above according to the results obtained by using them on the various benchmarks.

A. *Image Processing - non Machine Learning Techniques*

The first approach to the estimation of age and gender from face images was using image processing methods, and did not involve any Machine Learning, The method of using wrinkle extraction and contour smoothness in [8] worked well for gender estimation with an accuracy of 0.990 but did not work well for age estimation, which has an accuracy of 0.27. The primary reason for this would be the use of makeup of faces, which conceals wrinkles, and the difference in number of wrinkles for male and female faces.

B. *Linear SVM*

The linear SVM approach of [1], trained and tested on multiple benchmarks. The best results were obtained with the galaghar data-set and testing with the same with around 88.4%. The results are further given as part of Fig. 5.

C. *Basic Convolutional Neural Network*

A deep Convolutional Neural Network(CNN) is trained on face images to output the gender and label. The architecture in [2] proposed a 3 convolution blocks with varying number of filters and filter sizes. The results obtained for gender was around 86.8% with over sampling of the adience dataset and for age with 1-off error was around 84.7%. These results seem slightly less accurate than SVM from [1] but the advantage of CNN's or rather any deep learning method involves the accuracy to depict a generalised form as long as over fitting is avoided.

D. *Light-Weight Convolutional Neural Networks*

- MobileNet approach 1: Both age and gender predicted in the same output layer. The training for this was done in a very powerful machine with the following specs, Intel Xeon-E5, Nvidia GTX Titan X and 64 GB of RAM. The training on this machine took about 6 hours, this is due to the large dataset that's generally used for mobileNets this is to avoid under-fitting. Under fitting is a problem in mobileNets as the models are simple and small as compared to regular CNN's. MobileNets do not offer the same accuracy standards set by CNN's but they do offer significant improvement with respect to the time it takes per image. The benchmarks for time taken by this method of MobileNets is given in [Fig. 6]

- MobileNet approach 2: Age and gender predicted in different output layers. In [9] the experiments are run on the MacBook 2016 Pro laptop (CPU: 4xCore i7 2.2 GHz, RAM: 16 GB) and two mobile phones, namely:
1) Honor 6C Pro (CPU: MT6750 4x1 GHz and 4x2.5 GHz, RAM: 3 GB); and 3) Samsung S9+ (CPU: 4x2.7 GHz Mongoose M3 and 4x1.8 GHz Cortex-A55. The size of the model, and average inferencing time are shown in Fig3. The MobileNets are several times faster

than the deeper convolutional networks and require less memory to store their weights. Though the computing time for the laptop is signicantly lower when compared to the inference on mobile phones, their modern models (Mobile phone 2) became all the more suitable for oine image recognition. In fact, this model requires only

60 ms to extract facial identity features and predict both age and gender, which makes it possible to run complex analytics of facial albums on device. The highest gender recognition accuracy obtained is 0.99, on the Indian Movie Face Database(IMFDB)[9]. The highest age prediction accuracy obtained is 0.94 on the Eurecom Kinect Database [9], and 0.93 on IMFDB.

Fig. 4. The model size and inference time of the MobileNet developed in [9] and comparison with other models

CNN	Model size, MB	Average CPU inference time, s.		
		Laptop	Mobile phone 1	Mobile phone 2
age_net/gender_net	43.75	0.091	1.082	0.224
DEX	513.82	0.21	2.730	0.745
Our MobileNet	13.48	0.021	0.354	0.069
Our MobileNet, quantized	3.41	0.019	0.388	0.061

Fig. 5. Results of [1] with different training and testing benchmarks

Test	Train Source			
	Glgr	Glgr-Fnt	Ad	Ad-Fnt
Glgr	87.5 ± 0.2	87.1 ± 0.2	83.0 ± 0.3	82.9 ± 0.3
Glgr-Fnt	87.9 ± 0.2	88.4 ± 0.2	84.0 ± 0.2	84.9 ± 0.2
Ad	75.9 ± 1.2	75.8 ± 1.6	76.1 ± 0.9	75.2 ± 1.0
Ad-Fnt	77.4 ± 1.2	77.8 ± 1.3	77.2 ± 0.7	77.0 ± 1.0

Fig. 6. The speed of each model executed on mobile devices

Methods	Asus Zenbo Speed (ms/frame)	Asus Zenfone 3 Speed (ms/frame)
Levi_Hassner CNN [8]	≈ 4800	≈ 4800
LMTCNN-1-1	204.7	204.9
LMTCNN-2-1	297.6	367.2

V. CONCLUSIONS

Age and Gender estimation is a problem that is considered to be at the same difficulty level of face identification. How- ever, even as face identification and recognition has improved by leaps and bounds over the years, the former problem has not been dealt with as much. But with recent applications such as artificial personal assistants, album generation from galleries and face recognition for security purposes, the problem of age and gender estimation has gained importance, and more research is focused towards it. The solutions to this problem have transformed through the ages, starting from Image Processing based approaches, linear SVMs to deep neural networks and finally to light-weight Convolutional networks all the while focusing of improving the accuracy and reducing the inferencing time.

REFERENCES

[1] Age and Gender Estimation of Unfiltered Faces Eran Eifinger, Roee Enbar & Tal Hassner
[2] Age And Gender Classification Using CNN Gil Levi & TalHassner
[3] Joint Estimation of Age and Gender from Unconstrained Face Images Jia-Hong Lee, Yi-Ming Chan, Ting-Yen Chen, and Chu-Song Chen
[4] MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco An- dreetto, Hartwig Adam
[5] Implementation of Training Convolutional Neural Networks Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang, Jun Zhang
[6] Trainable Convolution Filters and Their Application to Face Recogni- tion Ritwik Kumar , Arunava Banerjee , Baba C. Vemuri , Hanspeter Pfister
[7] Human face anthropometric measurements using consumer depth camera Gargi Kabirdas Alavani , Venkatesh Kamat
[8] Age and Gender Estimation based on Wrinkle Texture and Color of Facial Images Jun-ichiro Hayashi, Mamoru YAUMOTO ; Hideaki ITO , Hiroyasu KOSHIMIZU
[9] Efficient Facial Representations for Age, Gender and Identity Recog- nition in Organizing Photo Albums using Multi-output CNN Andrey V. Savchenko