Illinois Institute of Technology

Project Report

Gender and Age Classification using Facial Features

Submitted by

Harsh Parikh
hparik11@hawk.iit.edu

Supervisor:

Prof. DR. Aron Culotta

Abstract

The principal objective of this project is to develop methods for the estimation of the gender and the age of a person based on a facial image. The extracted information can be useful in, for example, security or commercial applications. This is a difficult estimation problem, since the only information we have is the image, that is, the looks of the person. The training process needs to be optimized in terms of pre-processing, feature selection, choice of classifier and parameters. Automatic age and gender classification has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. Nevertheless, performance of existing methods on real-world images is still significantly lacking, especially when compared to the tremendous leaps in performance recently reported for the related task of face recognition. In this paper we show that by learning representations through the use of deep-convolutional neural networks (CNN), a significant increase in performance can be obtained on these tasks.

1. Introduction

Over the last decade, the rate of image uploads to the Internet has grown at a nearly exponential rate. Some of the most basic classifications regarding humans are gender and age. They are among the very first things a person decides on sight of someone. These decisions are based on many different features strictly coming from the person, but also from the environment. The marketing or sales departments of companies are usually interested in their products' targeted customers. Thus, it is important in many fields to have statistics of the target audience of their products. In the same way, some services, permissions or products are only allowed to an audience of certain gender or age, and it has to be somehow controlled. Bringing together the two ideas mentioned above, the need of estimating the gender and age in some automatic way appears. A human can easily make these estimates from faces. Yet, it is still a challenging task for a computer. This project is focused on gender and age estimation based on face images using computer vision techniques Applications for these systems include everything from suggesting who to "tag" in Facebook photos to pedestrian detection in self-driving cars. However, the next major step to take building off of this work is to ask not only how many faces are in a picture and where they are, but also what characteristics do those faces have. The goal of this project do exactly that by attempting to classify the age and gender of the faces in an image. Social media websites like Facebook could use the information about the age and gender of the people to better infer the context of the image. For example, if a picture contains many people studying together, Facebook might be able to caption the scene with "study session." However, if it can also detect that the people are all men in their early 20s and that some are wearing shirts with the same letters, it may predict "College students in a fraternity studying." Age and gender classification is an inherently challenging problem though, more so than many other tasks in computer vision. The main reason for this discrepancy in difficulty lies in the nature of the data that is needed to train these types of systems.

2. Related Work

[1]The areas of age and gender classification have been studied for decades. Various different approaches have been taken over the years to tackle this problem, with varying levels of success.

Early methods for age estimation are based on calculating ratios between different measurements of facial features. Once facial features (e.g. eyes, nose, mouth, chin, etc.) are localized and their sizes and distances measured, ratios between them are calculated and used for classifying the face into different age categories according to hand-crafted rules.

In this study, similar approaches are developed for gender and age estimation [7] which could be exploited to develop a more general system that can perform both tasks. First, face and eye detection are performed on the input image. After detection, alignment based on eye coordinates of the detected face image is applied to scale and translate face and reduces in feature space. Aligned face image is divided into local blocks and discrete cosine transform is performed on these local blocks. Concatenating features of each block, an overall feature vector is obtained. In gender estimation [3], SVM classifier is used for binary classification between female and male. In age estimation, a two-step classifier is used. In the first step, SVM classifier is used to discriminate between youth and adult and in the second step, support vector regression (SVR) is used for youth, adult and global age estimation to determine the specific age.

In recent years, with the dawn of never-before seen fast and cheap compute, revived the interest in [6] CNNs showing that deep architectures are now both feasible and effective, and continued to increase the depth of such networks to show even better performance. They advocate for a relatively shallow network, however, in order to prevent over-fitting the relatively small dataset they were operating on. Deeper networks, although generally more expressive, also have a greater tendency to fit noise in the data. So while improved performance with deeper architectures training on millions of images, shows improvements for shallower architectures for their use case [2].

3. Dataset

We collected data from LinkedIn profiles. Initially took 4 different LinkedIn IDs and recursively get other profiles based on LinkedIn recommended id list. So one by one, we scraped profiles and predict age from his/her school year and find gender based on first name using data.gov website. Sometime, it may possible that we get same Male and Female name so we ignore those profiles. We have put some threshold value to discriminate two genders. For age, we hardly find people put age/birth date on LinkedIn. So to calculate age, we find profiles' school/bachelor year and if we get it then we can add +18 and that we consider it as an age of a person. Sometimes we don't get bachelor degree information then we search for Masters or higher degree and calculate age based on that.

So we have collected around 8000 LinkedIn profiles and out of which 3000 have all useful information i.e. age and gender. In this profiles, we have around 1448 Female and 1652 Male. Most of the profiles' age fall between 20-35. Each image is annotated with the person's gender and age-group. The images are subject to various levels of occlusion, lighting, and blur, which reflects real-world circumstances. I used those images which were mostly front facing, which limited the total number of images to around 2,640. The images were originally of size 400x400, so they were pre-processed by all being resized down to 100x100.
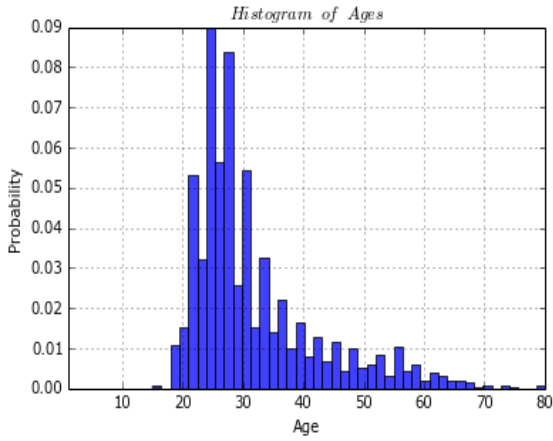
Fig 1. Histogram of profiles' age range

Fig 2. image dataset examples. Top row: 5 males of various ages. Bottom row: 5 females of various ages.

4. Method

The first thing to do while facing a big problem is to analyse it and divide it into different specific parts. A generic problem can be split into many steps with specific functions, so a tedious work is divided into small and well-defined parts.

Fig 3. Steps for Classification

An RGB image being input to the network is first scaled to 3x400x400 and then resized to 3x100x100. Then we have converted RGB image into Grey Scale image. So input size would be 100*100.

There are 3 convolution layers, followed by 3 fully connected layers.

4.1 Logistic Regression

Logistic regression is a probabilistic, linear classifier. It is parametrized by a weight matrix W and a bias vector b. Classification is done by projecting an input vector onto a set of hyperplanes, each of which corresponds to a class. The distance from the input to a hyperplane reflects the probability that the input is a member of the corresponding class. Mathematically, the probability that an input vector x is a member of a class i, a value of a stochastic variable Y, can be written as:

$$P(Y = i|x, W, b) = softmax_i(Wx + b)$$
$$= \frac{e^{W_i x + b_i}}{\sum_j e^{W_j x + b_j}}$$

The model's prediction y_{pred} is the class whose probability is maximal, specifically:

$$y_{pred} = \operatorname{argmax}_i P(Y = i|x, W, b)$$

```
Accuracy from inbuilt Method:  0.703579418345
            precision    recall  f1-score   support

         0       0.68      0.64      0.66      1192
         1       0.72      0.76      0.74      1490

avg / total       0.70      0.70      0.70      2682

[[ 759  433]
 [ 365 1125]]
```

Fig 4. Result of Logistic Regression for Gender

```
Accuracy from inbuilt Method:  0.578615779486
              precision    recall  f1-score   support

           0       0.65      0.73      0.69      1685
           1       0.42      0.34      0.38       987

avg / total       0.57      0.58      0.57      2672

[[1226  459]
 [ 652  335]]
```

Fig 5. Result of Logistic Regression for Age

4.2 Convolution Neural Networks (ConvNets) [9]

A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. [10] The architecture of a CNN is designed to take advantage of the 2D structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. In this article we will discuss the architecture of a CNN and the back propagation algorithm to compute the gradient with respect to the parameters of the model in order to use gradient based optimization.
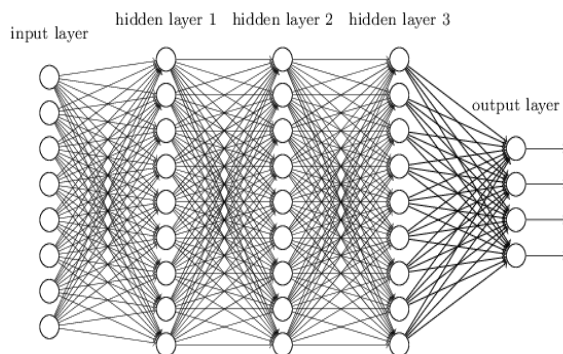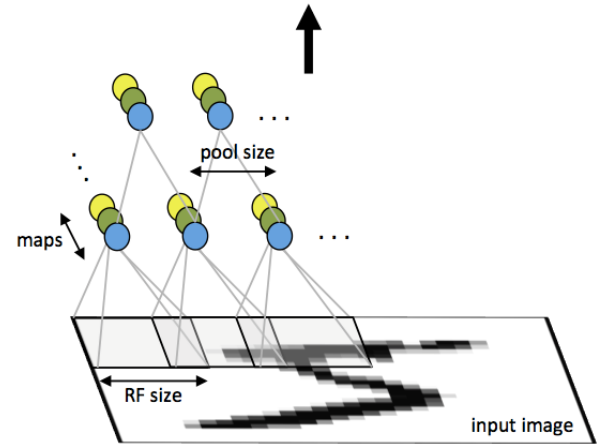

Fig 6. General Architecture of CNN

Fig 7. First layer of a convolutional neural network with pooling.

A CNN consists of a number of convolutional and subsampling layers optionally followed by fully connected layers [4]. The input to a convolutional layer is a m x m x r image where m is the height and width of the image and r is the number of channels, e.g. an RGB image has r=3. The convolutional layer will have k filters (or kernels) of size n x n x q where n is smaller than the dimension of the image and q can either be the same as the number of channels r or smaller and may vary for each kernel. The size of the filters gives rise to the locally connected structure which are each convolved with the image to produce k feature maps of size m−n+1. [8] After the convolutional layers there may be any number of fully connected layers. The densely connected layers are identical to the layers in a standard multilayer neural network.
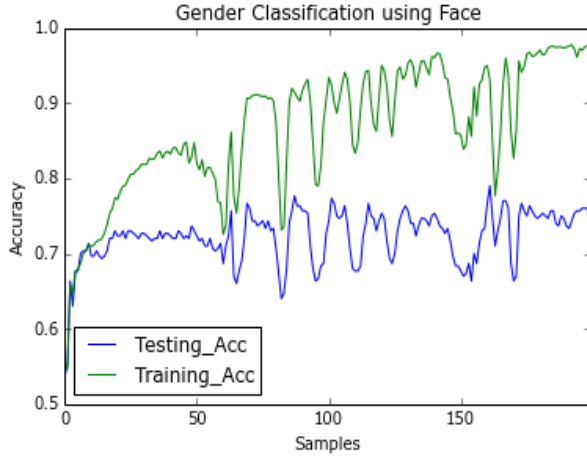
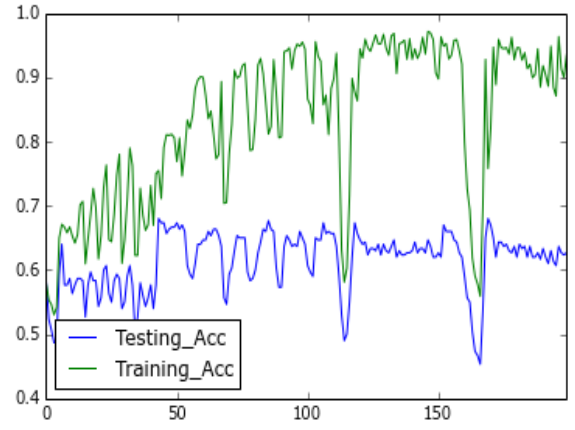*Fig 8. Training and Testing Accuracy of CNN for gender using face on 200 epochs (Accuracy vs Number of Samples)*

Fig 10. Training and Testing Accuracy of CNN for age using face on 200 epochs (Accuracy vs Number of Samples)
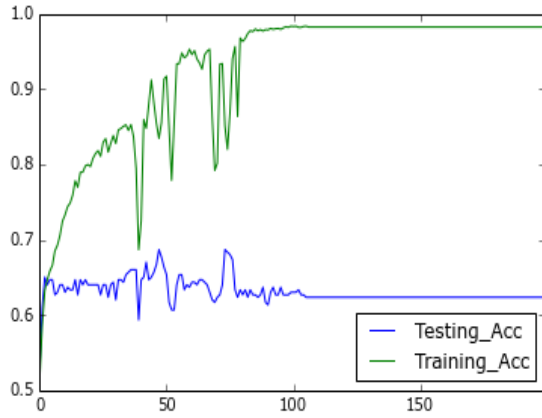

Fig 9. Training and Testing Accuracy of CNN for gender using whole image on 200 epochs (Accuracy vs Number of Samples)
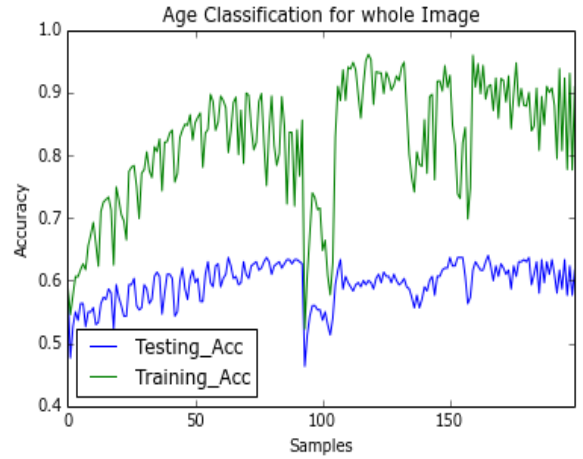

Fig 11. Training and Testing Accuracy of CNN for age using whole image on 200 epochs (Accuracy vs Number of Samples)

We have tested our model differently by tuning model parameters and sometimes changing network architecture and out of that we got training accuracy of 99.79% and testing accuracy of 81%. As in figure 8 shown, graph changes abruptly after few epoch cycles. Whereas in figure 9, we get stagnant accuracy results after 100 epochs for whole body image.

For Age estimation, we are getting training accuracy of 96.78%, but 68% testing accuracy only. There are certain reasons for that. Our dataset contains maximum 20-35 age group people. And for other age group we don't have much training data because LinkedIn users are mostly fall in this age group.
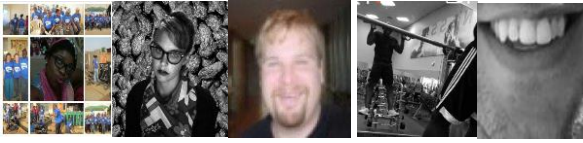
4.3 Misclassification Results

Figure 12. Gender misclassifications. Female subjects mistakenly classified as males and Male subjects mistakenly classified as females

Figure 13. Age misclassifications. Older subjects mistakenly classified as younger and Younger subjects mistakenly classified as older.

These show that many of the mistakes made by our system are due to extremely challenging viewing conditions of some of the Adience benchmark images. Most notable are mistakes caused by blur or low resolution and wrong working of face detection algorithm. Gender estimation mistakes also frequently occur for images where obvious gender attributes are not yet visible.

5. Conclusions

Although many previous methods have tackled the problem of age and gender classification of images, such settings do not adequately reflect appearance variations common to the real-world images in social websites and online repositories. I have used Adam Stochastic Gradient Descent Algorithm, but I will try another optimization algorithm as well. If there had been more time, I would have dedicated more effort towards fine-tuning the parameters and the modified architectures I experimented with. By far the most difficult portion of this project was collecting data, setting up the training infrastructure to properly divide the data into folds, train each classifier, cross-validate, and combine the resulting classifiers into a test-ready classifier. I foresee future directions building off of this work to include using gender and age classification to aid face recognition, improve experiences with photos on social media, and much more. Finally, I hope that additional training data will become available with time for the task of age and gender classification which will allow successful techniques from other types of classification with huge datasets to be applied to this area as well. I would try to remove the noisy data from my training data and reduce the features using PCA (Principle Component Analysis) and will use dropout. For age classification, I will classify into more age groups and try to get better accuracy.

References

[1] R. Martin-Felez, R. A. Mollineda, and J. S. Sanchez, "A Gender Recognition Experiment on the CASIA Gait Database Dealing with its Imbalanced Nature", in *International Conference on Computer Vision Theory and Applications VISAPP 2010), 2010, pp. 439444.*

[2] G. Mallikarjuna Rao, G. R. Babu,, G. Vijaya Kumari and N.Krishna Chaitanya, " Methodological Approach for Machine based Expression and Gender Classification," *IEEE International Advance Computing Conference,* pp. 1369-1374, 6-7 March 2009.

[3] T. Kobayashi and N. Otsu, "Image feature extraction using gradient local autocorrelations," *National Institute of Advanced Industrial Science and Technology,* 2008.

[4] Unsang Park, Yiying Tong and Anil K. Jain, "Face Recognition with Temporal Invariance: A 3D Aging Model," *Eighth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 1-7*, September 2008.

[5] Alice J. O'Toole, P. Jonathon Phillips, Fang Jiang Ayyad and Nils Pinard Herve Abdi, "Face Recognition Algorithms Surpass Humans Matching Faces over Changes in Illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1-18, September 2007.

[6] M. D. Malkauthekar and S. D. Sapkal, "Experimental Analysis of Classification of Facial Images," *IEEE International Advance Computing Conference*, pp.1093-1098, 6-7 March 2009.

[7] Ryotatsu Iga, Kyoko Izumi, Hisanori Hayashi, Gentaro Fukano and Testsuya Ohtani, "Gender and Age Estimation from Face Images," *International Conference on The Society of Instrument and Control Engineering*, pp. 756-761, August, 2003.

[8] Y. H. Kwon and N. da Vitoria Lobo. "Age classification from facial images. In Computer Vision and Pattern Recognition", *1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pages 762–767, Jun 1994.*

[9] http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/

[10] http://deeplearning.net/tutorial/lenet.html