Image based age and gender estimation

Đorđe Sević, SW-24/2016

1. Motivation

In the world where face recognition has become an everyday thing, defining some attributes of the recognized faces should not make a big deal. However, a growing interest has been emerging in age estimation from face images because of the wide range of potential implementations in law enforcement, security control, and human computer interaction. In these terms, recognition can still be a challenging issue, especially with low quality or non-aligned images.

The objective of this experiment is to improve the performance of the both age and gender estimation (classification). As the author said in [4], successes were made in the age estimation part, but the gender classification did not see any significant gains. Although the authors of [5] claimed that any deeper network would suffer from over-fitting, as it was cited in [4], in this experiment there will be some modifications on proposed CNN structure and values of parameters used for dropout and local response normalization.

2. Research questions

As it has been said, there are pretty decent systems whose main functionality is face recognition. However, different imaging conditions can lower their effectiveness. This project will try to face these problems using a proper dataset which can challenge these issues.

The dataset is provided by the Open University of Israel and it's OUI-Adience *Face Image Project*. The data included in this collection is 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. The sources of the images included in this set are Flickr albums. More information about dataset content is given in table 1.

Total number of photos	26,580
Total number of subjects	2,284
Age groups	8 [0-2] [4-6] [8-13] [15-20] [25-32] [38-43] [48-53] [60+]
Gender labels	Yes
In the wild	Yes
Subject labels	Yes

Table 1.

The dataset is composed of 5 folds to allow 5-fold 'leave one out' cross validation. To prevent overfitting, each fold contains different subjects. Each fold is described by a csv file with 12 columns, from which 4 are relevant to the given problem:

- *user_id* - the folder in the dataset containing the image.
- *original_image* - image name in the dataset.
- *age* - age label of the face.
- *gender* - gender label of the face.

The reason this type of protocol is necessary is because of the nature of the dataset being used, which contains multiple pictures of the same subjects. Therefore if the images were simply randomly shuffled and divided into fifths, the same subjects could potentially appear in both the training and test folds, thereby skewing the results to seem more promising than they are in reality.

This protocol therefore ensures that all the images of a given subject appear in a single fold to avoid this issue [4].

3. Related work

Very early attempts focused on the identification of manually tuned facial features and used differences in these features' dimensions and ratios as signs of varying age [4].

As early as 1990, neural networks were considered for the purposes of gender classification.

Later on in the early 2000s, support vector machines (SVMs) were used and it was found that they could achieve very low error rates on gender prediction of "thumbnail images" of subjects which were of very low resolution [4]. But in 2015, [5] broke this norm by developing one methodology and architecture to address both age and gender [4].

Its authors have been using the same dataset provided by the Open University of Israel. Their approach consists of several stages:

- *Face detection and alignment*
- *Encoded image representation*
- *Classification* - performed using standard linear SVM

In this one, and the several other cited papers, the face detection part is handled by Viola-Jones detection algorithm. Existing works mainly follow steps listed above, with the variations in the process of encoding image representation.

4. Methodology

Recently, CNN (Convolutional Neural Networks) model has proven to be the most suitable method for the classification task [3].

Network for this specific project is proposed to contain three convolutional layers, each followed by a rectified linear operation and pooling layer. The first two layers also follow normalization using local response normalization. The first convolutional Layer contains 96 filters of 7×7 pixels, the second convolutional Layer contains 256 filters of 5×5 pixels, The third and final convolutional Layer contains 384 filters of 3×3 pixels. Finally, two fully-connected layers are added, each containing 512 neurons [5].

As it has been said, after the first 2 pooling layers, there are local response normalization (LRN) layers. LRN is a technique that was first introduced in [7] as a way to help the generalization of deep CNN.

At the top of the proposed architecture lies a *softmax* layer, which computes the loss term that is optimized during training and also the class probabilities during a classification [4]. After we calculate the loss, we need to know how to minimize it in order to train an accurate classifier. The type of optimization used in this specific project is *stochastic gradient descent*.

5. Discussion

As it has been said while describing the dataset, training and testing will be performed using 5-fold cross validation with splits pre-selected to eliminate cases of images from the same Flickr album appearing in both training and testing sets in the same fold.

The performance of the two classifiers will be measured by the two standard metrics common in the literature: confusion matrix and accuracy. The confusion matrix is presented to the 8 classes (table 1.) age grouping results and for binary class gender classification results.

At first, the values of the hyper-parameters will be the same as those from [4], used in local response normalization. For the purpose of the experiment objective, there should be more effort towards fine-tuning the parameters and modifying proposed architecture.

6. References

[1] Eran Eidinger, Roee Enbar, and Tal Hassner, Age and Gender Estimation of Unfiltered Faces

[2] Abbas, A.R., Kareem, A.R.: Intelligent age estimation from facial images using machine learning techniques. Iraqi J. Sci. 59(2A), 724–732 (2018).

[3] Antipov, G., Baccouche, M., Berrani, S.A., Dugelay, J.L.: Effective training of convolutional neural networks for face-based gender and age prediction. Pattern Recogn. 72, 15–26

[4] Convolutional Neural Networks for Age and Gender Classification. Stanford University (2016)

[5] Levi, G., Hassncer, T.: Age and gender classification using convolutional neural networks. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, October 2015, pp. 34–42 (2015).

[6] Zhang, K., et al.: Age group and gender estimation in the wild with deep RoR architecture.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 1106–1114. 2012.