GeFace: Age and gender prediction from a single image of pure faces

Örkény Zováthi
Department of Control Engineering and Information Technology
Budapest University of Technology and Economics
o.zovathi@gmail.com

Máté Bártfai
Department of Control Engineering and Information Technology
Budapest University of Technology and Economics
matthaeusb@gmail.com

Abstract

In this paper we present a partly novel technique for age and gender estimation. Our approach combines well-known Convolutional Neural Network (CNN) architectures like Xception or SmallVGG for feauture extraction and prediction. The Network relies on pre-detected faces only. We believe that the face is a specific attribute and it contains all neccessary informations about the personality. Our goal was for this semester to see the efficience of this state-of-the-art methods for such images – like the face – where only small differences egsists. We consider the features (age, gender) as two undependent tasks, therefore we created own branches for each task and trained them separately as well. At the outputs, we considered the gender as a binary classification and the age as a regression problem. We tested our network qualitative and quantitative and the results showed that in most cases the gender and the age group can be pretty well predicted. In the future we plan to fine-grade the network and/or combine it with other techniques to make it useful for more responsible tasks (e.g. age calculation for people with undocumented birth), not just for fun.

1 Introduction

In the current world, *deep learning* techniques and *Artificial Neural Networks* became more and more popular[4]. These systems can „learn" to perform tasks by considering examples, generally without being programmed with any problem-specific rules. Practically it means that we do not have to own a deep knowledge of the field of application (e.g. natural language processing, image analysis): if we constuct a „good" deep neural network, we are still able to produce such systems that learns and applies the background dependencies and consequenses[10].

In the field of image analysis, age estimation is an open and unsolved task in today's life. Although, in the past few years a lot of different approaches were created and presented, we are still far away from the perfect result. Some of these previous methods collect a lot of information about the person (e.g. height, weight, favourites, family status), other approaches look the whole or upper body to gather their inputs.

Our proposal and motivation was to predict the correct age from a single image of the face, with no additional information. It is probably not the most efficient approach but as we have already mentioned, we focused on the face as an individual and specific attribute. With the help of Convolu-

Deep Learning in Practise with Python and LUA (2018), Budapest.

tional Neural Networks (CNN), our goal was to learn the facial features[9] and combining the age prediction with a gender estimation in order to create a primitive one-shot person analyser system what we call GeFace.

Here we want to notice that the project do not end at the end of the semester, we want to improve it with other techniques and methods as well. But, in this first round, our goal was to see whether such well-known architectures that reached a good result on the *ImageNet*-challenge are also able to produce „good” results for face images, where the feature-differences between the images are hardly noticeable.

2 Pipeline

Our constructed intelligent system consists of a built-in OpenCV FaceDetector[2], a Preprocessing part, a two-branch Deep Neural Network and a visualizer deployment part. The pipeline can be seen in Figure 1.

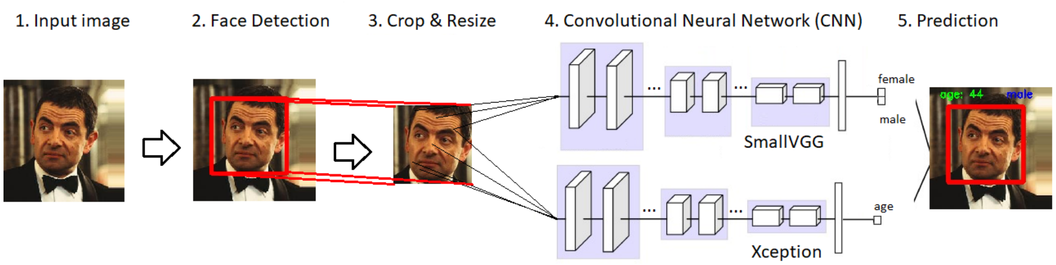


Figure 1: The pipeline of our approach. We used a two-branch network for age and gender prediction.

In the following subsections we will describe each part in details.

2.1 Dataset

For labelled input images, we used the IMDB Dataset[5],[6] which contains around 500 000 pictures of human upper body. The metadata of the dataset was shared as a *.mat* file so we transformed it to *.csv*. This new data contains only the image path, the age and gender value, and the file can be directly processed by Python and Pandas as well. Moreover, after we looked into some pictures and analysed them we found out that in a lot of cases the faces on the pictures are unavailable to recognise or are too small for evaluation purpose. To drop the unusable pictures and crop the faces for laterer use, we used the built-in OpenCV FaceDetector and runned it for all of the images in the dataset. On every picture where a face could be recognised, we cropped it out, resized it to $96 \times 96 \times 3$ and saved it back to our new database. As result, we got ca. 250 000 face images. We made the formed dataset available[1] on the following link:

`https://drive.google.com/open?id=14I8YEH0egjkkbrpaBIAcQuoM9cGHVUoH`

The size of the dataset is about 2.5 GB and the metadata is also attached. [1]

3 Preprocessing and standardscaling

Since we used the OpenCV FaceDetector for creating the database, we also had to apply it as a preprocessing step for predicting and deploying. We also splitted our dataset into train, validation and test parts. Our rule was 70-10-20. We experienced that the structures of the available face images are quite homogen and similar, so we scaled the input image intensity values for all pixels into $[-1, 1]$ in the following way:

$$I_{new} = I_{old} * \frac{2.0}{255.0} - 1$$

[1]Please contact one of us in the given e-mail addresses, if not!

For the output values of the network, we used a MinMaxScaler for the ages and a LabelBinarizer – which works like the OneHotEncoder – for the gender values. These normalizers were fitted on the training images but we transformed all three datasets (train, validation and test). We wrote out the instances (Binarizer, MinMaxScaler) to files for the laterer deployment phase as well.

4 Network architecture

As we have already said, the goal of this milestone was to try out well-known architectures for the face analysis. We considered the two outputs undependent so we created two branch[3] from the Input layer. For age estimation we used the Xception[1] network architecture, which ownes one of the best results on the *ImageNet*-challenge. We did not load any pre-trained weights and exluded the top layers as well. We extended the branch with a *Dropout*, a *Flatten* and a *Dense* layer with only one output at the end which is the scaled value of the predicted age.

For gender prediction, we constructed an other batch by ourself based on the VGG-16 architecture[7]. From the number of layers we called it as a SmallVGG branch. The parameter space of the branches is the following:

Xception branch:	21 million
SmallVGG branch:	0.5 million

It can be seen that we used with two order of magnitude more parameter for the age as for gender. This ratio is similar as the ratio between the possible outcomes of the two branches.

5 Training

We trained our final model on a TitanX GPU around $4-5$ hours. We splitted the training process into two parts: First, we disabled the trainability of every layer of the *SmallVGG* branch and trained only the *Xception* layers, secondly, we inverted the trainability and trained only the SmallVGG layers. We also applied a prior knowledge there, since we experienced that the SmallVGG needs more epoch for the training than the Xception. The reason can be for this that we applied many *Dropout* layers for avoiding the overfitting.

We used the Adam optimizer for backpropagation. During the training, we monitored the output loss for the age part, and the accuracy for the gender fine tuning.

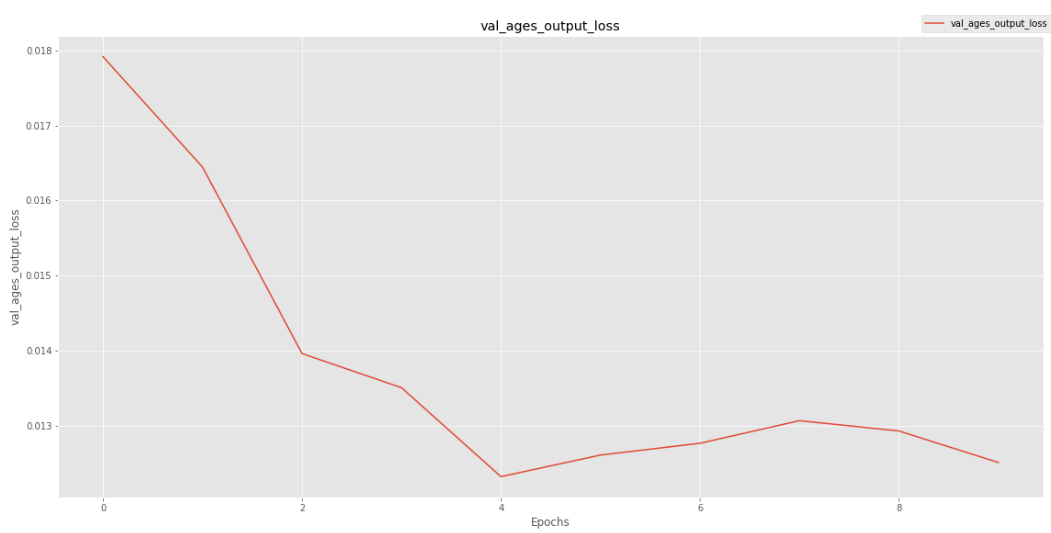


Figure 2: Validation age loss.

As you may noticed from the figures, the training do not stopped on the best result. We applied Earlystopping and CheckPoint to save the best model, but also allowed a 5 epochs patience.

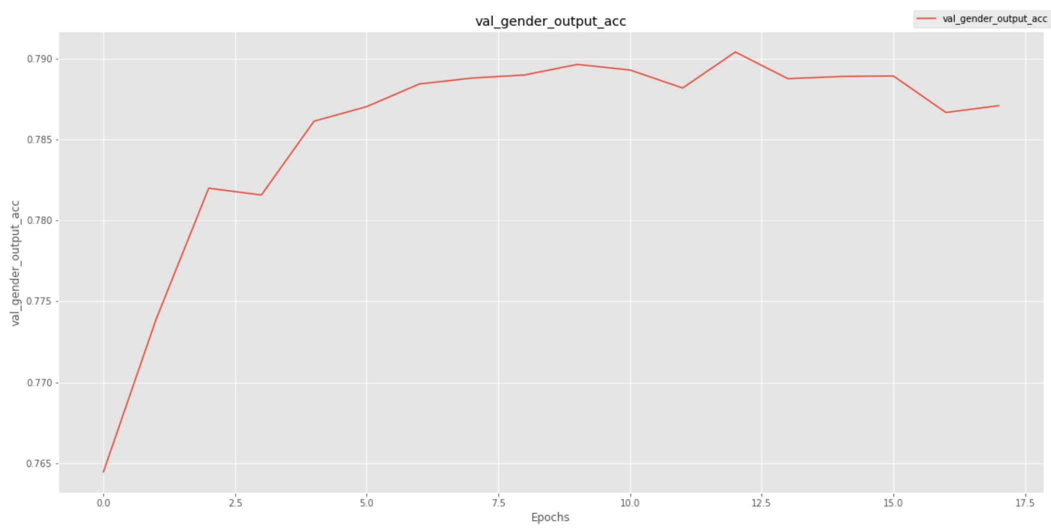


Figure 3: Validation gender accuracy.

Hiperparameters

We trained batch-wise with a batch size of 64. We also applied a decreasing learning rate with an initial value of 0.001. In every epoch, the actual learning rate was the inital value divided with the number of the epoch. We also tried out other hiperparameter settings but they showed worse results.

6 Testing

After we successfully trained and fine-tuned the network, we tested the prediction results qualitative and quantitative. Our test database contained 25059 face images from the IMDB dataset. We created a statistic analysis about the age regression and a confusion matrix for the gender prediction. To analyse the results of the predicted ages, we used statistical methods:

Average distance:	8.2 years
Variance:	7.5 years

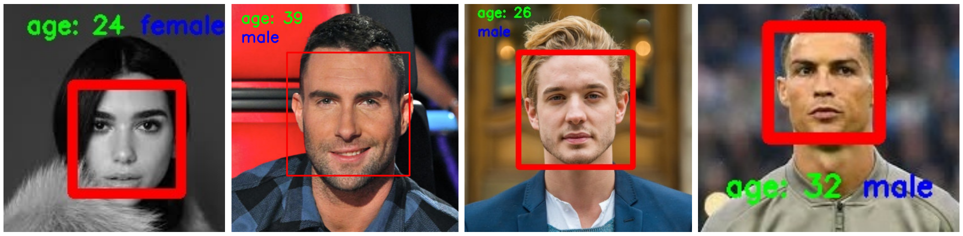
For the gender prediction, we used the standard classification metrics:

	Precision	Recall
Man:	0.81	0.8
Woman:	0.76	0.78

The results relies on the following *confusion matrix*, which was constucted after the whole test dataset.

Confusion matrix	Man	Woman
Man	11366	2867
Woman	2597	9229

We also checked the output of our system for specific cases to see how it works in practise. In conclusion, for pictures with similar structure as the IMDB dataset (e.g faces of stars, famous people) the network worked pretty well. Of course none of these pictures was in the training data. Some of the good results are shown in Figure 4. In the most testcases the age group and the gender was correctly predicted. We have also compared the network with human predictions. The results showed that the system is able to predict the same level as we did it.



(a) Gender: Female
Current age: 23
Predicted age: 24

(b) Gender: Male
Current age: 39
Predicted age: 39

(c) Gender: Male
Current age: 26
Predicted age: 26

(d) Gender: Male
Current age: 33
Predicted age: 32

Figure 4: The model's prediction and the ground truth for a few „good" test cases.

(a) Gender: Female
Current age: 34
Predicted age: 28

(b) Gender: Female
Current age: 49
Predicted age: 35

(c) Gender: Male
Current age: 47
Predicted age: 38

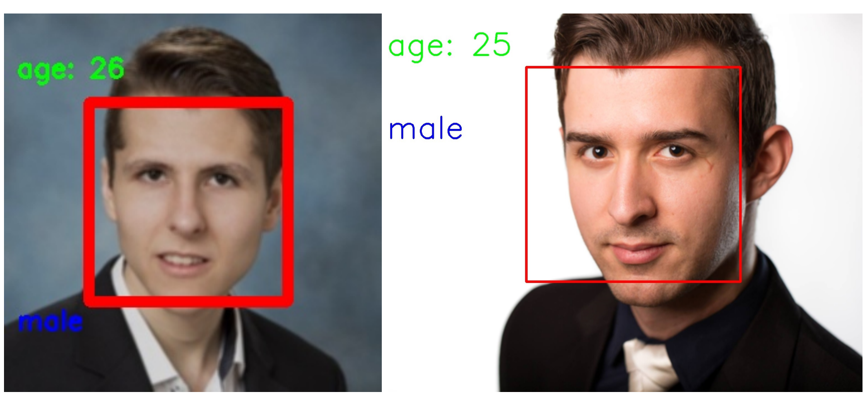
(d) Gender: Female
Current age: 46
Predicted age: 31

Figure 5: The model's prediction and the ground truth for a few „worse" test cases.

7 Deployment

In the first round we created a Jupyter Notebook for the deployment to use the model for fun. In the near future we plane to make it „online" and with the contribution and allowance of the users we want to collect an own dataset for future purpose. The code and implementation is available on GitHub:

`https://github.com/bartfaimate/GeFace`



(a) Gender: Male
Current age: 24
Predicted age: 26

(b) Gender: Male
Current age: 23
Predicted age: 25

Figure 6: The model's prediction for us. We were not in the training database of course.

8 Conclusion

During the semester we have got to know how deep learning techniques work in practise. We constructed a network which predicts age and gender from a single image of the face. The trained network was able to predict the age group and gender in most cases. When we tested the results with us, the network showed us in average a few years older. The reason behind this effect is that our training dataset contained actors and actress with beautified faces and we not.

Since this approach is a hard way for the best efficience we want to try out other network architectures for the same task as well. For example, [8] uses local informations for car brand predictor and it can be also useful for face images. We also plan in the future to combine our results with other networks.

References

[1] François Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv e-prints*, page arXiv:1610.02357, October 2016.

[2] Fares Jalled and Ilia Voronkov. Object Detection using Image Processing. *arXiv e-prints*, page arXiv:1611.07791, November 2016.

[3] Wenhan Liu, Qijun Huang, Sheng Chang, Hao Wang, and Jin He. Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram. *Biomedical Signal Processing and Control*, 45:22 – 32, 2018.

[4] Nicholas G. Polson and Vadim O. Sokolov. Deep Learning. *arXiv e-prints*, page arXiv:1807.07987, July 2018.

[5] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, December 2015.

[6] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*, July 2016.

[7] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, page arXiv:1409.1556, September 2014.

[8] Yaming Wang, Vlad I. Morariu, and Larry S. Davis. Learning a Discriminative Filter Bank within a CNN for Fine-grained Recognition. *arXiv e-prints*, page arXiv:1611.09932, November 2016.

[9] Chenjing Yan, Congyan Lang, Tao Wang, Xuetao Du, and Chen Zhang. Age estimation based on convolutional neural network. In Wei Tsang Ooi, Cees G. M. Snoek, Hung Khoon Tan, Chin-Kuan Ho, Benoit Huet, and Chong-Wah Ngo, editors, *Advances in Multimedia Information Processing – PCM 2014*, pages 211–220, Cham, 2014. Springer International Publishing.

[10] Rong Zhang, Weiping Li, and Tong Mo. Review of Deep Learning. *arXiv e-prints*, page arXiv:1804.01653, April 2018.