# Age and Gender Classification for Precision Marketing

Ouyu Lan and Jiaxian Guo

***Abstract* — The information of age and gender is important for precision marketing. Therefore, our goal is to automatically recognize the age and gender from a still photo or active video. In this paper, we show that the deep convolutional neural networks, VGG-16, achieve a significant improvement in the field of age and gender classification in real-world. We train and evaluate our method on Adience benchmark and get a high accuracy.**

***Index Terms* — age classification; deep convolution neural network; gender classification; precision marketing**

## I. Introduction

Precision marketing is more and more important in the commercial society, which emphasizes relevance between the advertisements and the target audience. Age and gender are two of the most important factors to measure the correlation. However, the ability to predict the age and gender accurately and reliably from face photos and videos has not reached the commercial demand.

There has been a lot of approached to predict the age and gender from face images. In the earliest stage, the classification is based on difference in facial feature dimensions [4], or some specific algorithms. Afterwards, many machine learning methods are applied to the age and gender recognition system. Deep convolutional neural networks (CNN) is considered to be one of the best methods in image and video recognition. Recently, lots of attempt have been made to improve the original architecture of AlexNet to achieve better accuracy in the field of age and gender classification. For instance, Gil Levi and Tal Hassner [1] used a smaller number of neurons and convolution layers, which achieve a state-of-art accuracy in 2015.

In this paper, we attempt to improve the age and gender prediction accuracy to meet the commercial demand. To this end, we follow the successful models in image classification [2], which achieved the state-of-the-art accuracy on ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) in 2014. Relative to previous approaches, our model uses the VGG-16 architecture [2], which has a deeper network architecture and smaller convolution filters. We classify eight classes for age and two for gender, which are suitable for precise marketing.

Our network is pretrained on IMDB-WIKI dataset [3]. Then we train and test our network on the Adience benchmark, which is designed for age and gender classification without prior filtering [5]. That means these pictures can reflect the real-world face images without conditional control, such as lightness, sharpness, angles, etc..

Taking into account the windows operating system has the largest proportion of the world today and C# develop the windows software conveniently, so we develop a SDK for C# in Caffe to use Caffe model, this SDK can help any C# to develop windows application using Caffe model, It will help AI application developed in windows. And then we use it develop the application which play ads based on age and gender to realize précising marketing.

## II. Existing Work

Early methods are different for age and gender classification relatively. At first, the recognition is based on the size, distance and location of facial features. This method requires much extra information for age and gender classification and high face images quality. Later, with the development of machine learning, neural networks, especially CNNs, achieve much better performance in this field. [1] reaches the accuracy of 86.8% in gender classification (2 classes) and 50.7% in age classification (8 classes). [3] also gets good performance on its own dataset.

## III. Proposed Methodology

Our proposed method is to pre-process input images, extract their features through VGG-16 architecture and then predict the value. We implement it by Caffe open-source framework.

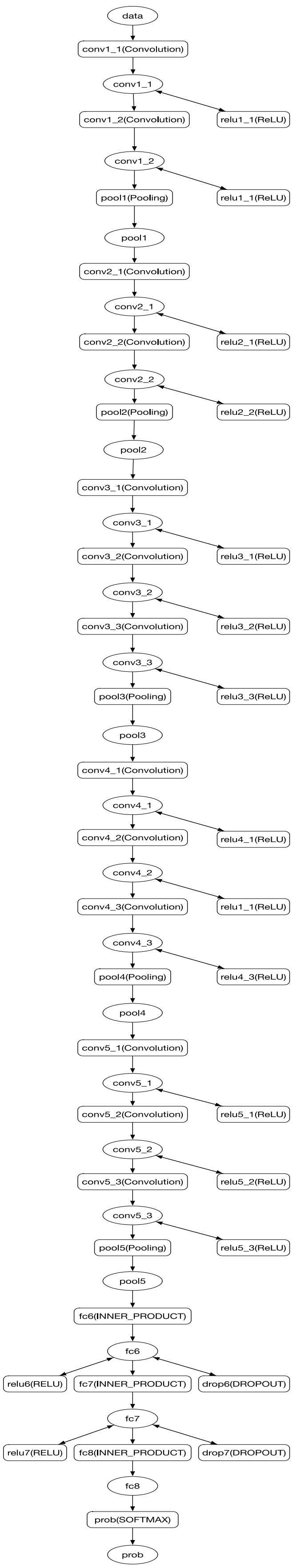


Figure 1. VGG-16 architecture.

### A. *Datasets.*

To get enough labeled image training set for age and gender estimation is a hard job. We cannot prove the truth of the photo information, such as the person's birth date, the date when the photo was taken and even the gender. Furthermore, we need real-world social images for our application, instead of altered photographs of celebrities or others posted on Internet.

| | 0-2 | 4-6 | 8-13 | 15-20 | 25-32 | 38-43 | 48-53 | 60- | Total |
|---|---|---|---|---|---|---|---|---|---|
| Male | 745 | 928 | 934 | 734 | 2308 | 1294 | 392 | 442 | 8192 |
| Female | 682 | 1234 | 1360 | 919 | 2589 | 1056 | 433 | 427 | 9411 |
| Total | 1427 | 2162 | 2294 | 1653 | 4897 | 2350 | 825 | 869 | 19487 |

Table 1. The number of images in different age and gender categories in Adience set.

The images in Adience set [5] are automatically uploaded to Flikr from mobile devices without any manual filtering or constrained conditions. It consists of nearly 26K images of 2,284 subjects. Table 1 shows the number of images in different age and gender categories [1]. It is evident that the images are not well-distributed. Considering the software audience group, we do not redistribute them. We divide it into three parts, training, validation and testing for both age and gender classification and compare the reported results in [1] to the results by our network.

### B. *Network Architecture.*

Our network uses the VGG-16 architecture [2] which gets impressive performance in image classification for both age and gender classification. It is illustrated in Figure 1. The network includes thirteen convolutional layers and three fully-connected layers. At the final layer, we classify the age into eight classes and gender into two classes, which can get a better performance when making precision marketing.

The image fed to the network are first resized into 256×256 and cropped into 224×224 pixels with three GRB channels. The first two convolutional layers (conv1_1 and conv1_2) contain 64 filters of size 3×3 pixels, followed by a max pooling layer in 2×2 regions. The third and fourth convolutional layers (conv2_1 and conv2_2) contain 128 filters and then follows by a max pooling layer. Then fifth to seventh convolutional layers (conv3_1, conv3_2 and conv3_3) contain 256 filters, eighth to tenth (conv4_1, conv4_2 and conv4_3) 512, eleventh to thirteenth (conv5_1, conv5_2 and conv5_3) 512. Each stage is followed by a max pooling layer and set the same hyper-parameters as before. Then follows three fully connected layers. The first receives the output of thirteenth convolutional layer with 4096 neurons. The second receives the output of the first fully connected layer with the same number of neurons. Both two layers are followed by a ReLU and a 50% dropout layer. The last fully connected layer maps the second fully connected layer to the classes for age and gender, whose output is fed to a softmax layer indicating the possibility for each class. Finally, we choose the maximal probability as the prediction.

### C. Training and Testing.

The weights in the network are fine-tuning from the pre-trained model of [3]. Therefore, we can get a better performance for our own application. Given a test image, we can get a vector of probability in the dimension of the number of class with 1 for ground truth and 0 elsewhere.

We apply two dropout layers with a ratio of 0.5, a random crop of 224×224 pixels to avoid the over-fitting. This is similar to [1]. During the training, we use stochastic gradient decent and decay learning rate

### D. Building SDK for C# in Caffe.

Caffe is a famous deep-learning tools in image process area, It greatly makes the scientists more convenient to do experiments. But it only has C++/Python/Matlab interface, these 3 languages is not suitable for engineering development. So we packaged some Caffe's interface for C#, and make it SDK, if you build a C# solution, you can use this SDK and then call your own Caffe model to begin your application, it greatly promoted the conversion of research to industry with three features:

1) *Base function:*

Input Image, Output classification and confidence (Resize Image automatically), Support multi-lable (i.e. Human face have age label and gender label)

2) *Convenience:*

Reference our SDK and Caffe model to develop a new application.

3) *Expandability:*

You can make any wrapper and add features or function above it.

### E. Precision marketing

For each age and gender category, we predefine the proper advertisement. Table 2 lists the recommendation.

| | 0-2 | 4-6 | 8-13 | 15-20 | 25-32 | 38-43 | 48-53 | 60- |
|---|---|---|---|---|---|---|---|---|
| Male | Baby formula | Brain-building toys | Brain-building toys | Keyboard | Keyboard | Car | Car | Health care |
| Female | Baby formula | Doll | Doll | Smartphone | Smartphone | Make up | Make up | Health care |

Table 2. The recommended advertisement for different categories.

## IV. Experimental Results and analysis

Table 3 shows the best results in [1] and our results for age and gender classification. It is apparent that our proposed method achieves considerable higher accuracy in both age and gender classification tasks. We can see that deeper CNN can have a better performance if we have enough data to train the model.

| classification | Method | Accuracy |
|---|---|---|
| Age | Best from [1] | 50.7 |
| | Proposed method | 69.0 |
| Gender | Best form [1] | 86.8 |
| | Proposed method | 95.2 |

Table 3. Age and gender estimate result on the Adience benchmark.

During testing, we find that the model does not such well in recognizing the age and gender of a face in a much larger background, i.e. the face region just takes a small part of the whole given image. Therefore, the step for face detection and capture is necessary.

## V. Conclusion and Future work

Our proposed method uses deep CNN with VGG-16 architecture pre-trained on IMDB-WIKI. And we can get enough high accuracy in both still photos or active frames of videos in real-world.

By capturing and analyzing the faces on the street or at the back seat of a taxi, we can play different advertisements which may get the attention of the people with the predicted age and gender.

## Acknowledgment

The research is supported by the course EE312 and CS362 Software Engineering in Shanghai Jiaotong University.

## References

[1] Gil Levi and Tal Hassner, Age and Gender Classification using Convolutional Neural Networks, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, June 2015.

[2] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2014.

[3] Rasmus Rothe and 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), 2017.7.

[4] Y.H. Kwon and N.da Vitoria Lobo. Age classification from facial images. In Proc. Conf. Comput. Vision Pattern Recognition, pages 7620767. IEEE, 1994.

[5] E. Eidinger, R. Enbar, and T.Hassner. Age and gender estimation of unfiltered faces. Trans. On Inform. Forensics and Security, 9(12), 2014.