Age Classification Using Convolutional Neural Networks with the Multi-class Focal Loss

Wei Liu¹, Lin Chen², and Yajun Chen³,*

¹School of Computer Science & Technology, China West Normal University, No.1, Shida Road, Shunqing District, Nanchong, China
²Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, No.266 Fangzheng Avenue, Beibei District, Chongqing, China
³Education and Information Technology Center, China West Normal University, No.1, Shida Road, Shunqing District, Nanchong, China

*Corresponding author: scncdyj@cwnu.edu.cn

Abstract. Automatic age classification has drawn significant interest in plenty of applications such as access control, human-computer interaction, law enforcement and surveillance. Automatic age classification is a challenging task due to the complexity of facial images. A large number of approaches have been investigated on unconstrained datasets. However, most of these approaches have focused on the network architecture rather than the distribution of data, i.e., the extreme class imbalance existing among different age groups as the difficulty of data collection. In this paper, we propose a convolutional neural networks model based on the multi-class focal loss function. Specifically, our approach is designed to address the class imbalance via reshaping the standard cross entropy loss that it down-weights the loss assigned to well-classified examples. We validate our approach on well-known Adience benchmark. Finally, the experimental analysis shows that the proposed model achieves a significant improvement in performance for age classification.

1. Introduction

Human face conveys many kinds of important information, such as identity, expression, emotion, gender, age, etc. With the rapid emergence of the intelligent applications, there is an increasing demand for automatic extraction of facial attributes. As one of the key facial attributes, age plays a fundamental role in social interactions. Therefore, automatic age estimation from facial images is an important and challenging task investigated in numerous applications, including access control, human-computer interaction (HCI), law enforcement and surveillance [1].

Automatic age estimation aims to assign a label to a face image regarding the exact age or the class of age it belongs. Most of the previous approaches for age estimation are intended for accurate estimation of the actual age. However, as a result of the speciality of aging effects on the face, it is difficult to estimate an actual age. There are different aging patterns among different people, which are influenced by many internal and external factors, including genes, ethnicity, health condition, lifestyle and environment [2]. Therefore, the aging progress is uncontrollable, making age estimation challenging for humans in some cases, and even greater challenging for computer vision systems [3]. Furthermore, the ability of automatic age estimation is influenced by additional factors, such as lighting conditions, facial expressions and pose variations.
In recent years, the automatic age classification of facial images has been investigated extensively[4]. In particular, convolutional neural networks (CNNs) have attracted substantial research attention [5, 6], which are capable of learning a compact and discriminative feature representation with large-scale data. As a consequence, CNNs achieved a great success in the vision community, significantly improving the performance in classification problems.

In this study, we concentrate on the issue of age group classification rather than that of exact age estimation. Though extensive approaches for age classification have been proposed, most of them focused on constrained images such as FG-NET [7] and MORPH [8]. Unlike these methods discussed above, we devote to unconstrained images and study the issue of extreme class imbalance among different age categories: there are a great number of face images in the age group of 25-32, while a few face images belong to the older groups. In this paper, we propose a CNN model based on the multi-class focal loss function, to efficiently improve the performance of age classification. Our main contributions are summarized as follows:

- We propose a CNN model based on the multi-class focal loss function to address the imbalance of different classes for age classification.
- Experimental results show that the proposed approach could benefit the class imbalance problem in the CNN based classification. Despite of the very challenging nature of the images in the Adience dataset, our approach achieves significant improvements in age classification accuracy over the state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 briefly reviews the related work. Section 3 illustrates the proposed method and network architecture in detail. The experimental results are given in Section 4. Finally, conclusions are drawn in Section 5.

2. Related work
A comprehensive review of age representation methods from facial images was presented in [9]. Kwon et al. [10] proposed a method to classify images into different age categories by calculating ratios of different measurements based on facial features. However, this method may not be suitable for in-the-wild images which have a large of variations in pose, illumination, expression and occlusion. Geng et al. [11] proposed an automatic age estimation method named AGES, which constructs a representative subspace and models the face pattern in time order as a sequence of the individual face images. However, this method demands for face images of a person at different ages. In order to address this problem, Fu et al. [12] proposed a novel manifold space that analyzes and learns the face manifold for different age stages to find a low-dimensional embedding space.

Feature extraction is a crucial step in automatic age estimation. Several methods for feature extraction have been proposed, for instance, Active Appearance Model (AAM) [13], Local Binary Patterns (LBP) [14-16], Anthropometric Features [17] and Biologically-Inspired Features (BIF) [18]. In addition, regression and classification methods are utilized to estimate the accurate age or the category of age from facial images. SVM was utilized for age classification in [19]. Some methods for regression were utilized to predict the accurate age, such as SVR [20], linear regression [21] and CCA [22]. Even though all of these methods have proven effective on constrained benchmarks, they could fail to tackle large variations in unconstrained images.

Recently, increasing attention has been investigated on CNN for age estimation from facial images. Yi et al. [23] proposed an age estimation method based on Multi-Scale CNN to learn parameters instead of hand-crafted. Levi et al. [5] proposed a simple CNN for age classification on unconstrained Adience benchmark, to effectively avoid the overfitting problem. Ranjan et al. [24] proposed an approach based on deep convolutional neural networks (DCNNs) for age estimation on unconstrained challenge dataset, and outperformed conventional methods. Residual Networks of Residual Networks (RoR) [25] was capable of tackling large variation in the wild for age classification. Chen et al. [26] proposed ranking-CNN via taking the ordinal relation between ages into consideration for age estimation.
To summarize, CNNs have achieved great success on age classification task, significantly improving the performance of classification. Most of the existing work improved classification accuracy via modifying the network architecture. However, the age estimation with imbalanced distribution has not been investigated in depth so far. In this study, we propose a CNN model based on the multi-class focal loss function, which achieves better results than state-of-the-art methods.

3. Proposed method
In this section, we elaborate our multi-class focal loss and describe the details of the network architecture.

3.1. Multi-class focal loss
The original focal loss [27] was designed to address the class imbalance for binary classification but it can be extended to address the multi-class classification problems. For multi-class issue, the Cross Entropy (CE) loss for an example is given in equation (1):

$$L_{CE} = -\sum_{i=1}^{C} t_i \log(y_i)$$  

(1)

where $C$ denotes the number of categories, $t_i$ denotes a real probability distribution, $y_i$ denotes a probability distribution of the prediction. As shown in equation (2), where $t_i = 1$ if $i$ belongs to the true label, else it is 0.

$$t_i = \begin{cases} 1 & (i = \text{true label}) \\ 0 & (i \neq \text{true label}) \end{cases}$$  

(2)

The Adience dataset is split into eight age categories. Moreover, there is a great difference in the number of different categories. It is difficult to classify the age groups of 15-20, 38-43, and 48-53. In contrast, the age groups of 0-2 and 25-32 are more easily classified.

Our multi-class focal loss is designed to address the class imbalance via down-weighting easy examples such that their contribution to the total loss is small even if their number is large. In other words, it focuses training on minority examples. A method for addressing class imbalance is to add a modulating factor $(1 - y_i)\gamma$ to the cross entropy loss, with adjustable focusing parameter $\gamma \geq 0$. We define the multi-class focal loss, as formulated in equation (3):

$$L_{fl} = -\sum_{i=1}^{C} (1 - y_i)^\gamma t_i \log(y_i)$$  

(3)

There are two parameters related to the multi-class focal loss. In the first place, a focusing parameter $\gamma$ is utilized to control the rate at which easy examples are down-weighted. Especially, when $\gamma = 0$, $L_{fl}$ is equivalent to $L_{CE}$. The effect of the modulating factor $(1 - y_i)^\gamma$ is increased with $\gamma$. Secondly, when an example is misclassified and $y_i$ is small, the modulating factor nearly tends to 1, and thus the loss is unaffected. As $y_i$ increases to 1, the modulating factor nearly tends to 0 and the loss for well-classified examples is down-weighted. We adopt the multi-class focal loss in our experiments with $\gamma = 1.5$, which achieve significant improvements in age classification accuracy.

3.2. Network architecture
Our experiments for age classification are implemented using CNN with the multi-class focal loss. As depicted in Figure 1, the network contains three convolutional layers and two fully connected layers. The choice of the simple model aims to reduce the risk of overfitting [28]. Unlike the network architecture proposed in [5], we adopt the multi-class focal loss function instead of softmax loss function to effectively address class imbalance. During training, the input image is resized to 256×256 pixels, and then a random crop of 227×227 pixels is fed to the network. Moreover, the image is passed through three convolutional layers as follows:
Figure 1. The CNN architecture for age classification.

The first two convolutional layers are followed by a rectified linear unit (ReLU) as the activation layer, a max pooling layer and a local response normalization layer [6] with the same hyper parameters, respectively. Then, the third convolutional layer is followed by only a ReLU and a max pooling layer. Moreover, all max pooling layers are taken the maximal value of 3×3 regions with two-pixel strides. Furthermore, spatial padding is utilized in the second and the third convolutional layers, to preserve the resolution after the convolution operation.

To reduce the risk of overfitting, dropout learning [29] is utilized after two fully connected layers. The remaining architecture of the network is as follows:

The configuration of two fully connected layers is consistent, which contains 512 neurons, followed by a ReLU and a dropout layer at a dropout ratio of 0.5. In the final layer, the output from the previous layer of size 512 features is densely mapped to 8 neurons for age classification. A soft-max with the multi-class focal loss layer is adopted to obtain a probability for each class. More details of the network architecture are given in Table 1.

Table 1. Detailed architecture of the CNN for age classification.

| Layer  | Kernel size | Stride | Pad | Output size |
|--------|-------------|--------|-----|-------------|
| Input  | -           | -      | -   | 227×227×3   |
| Conv1  | 7×7         | 4      | -   | 56×56×96    |
| Pool1  | 3×3         | 2      | -   | 28×28×96    |
| Conv2  | 5×5         | 1      | 2   | 28×28×256   |
| Pool2  | 3×3         | 2      | -   | 14×14×256   |
| Conv3  | 3×3         | 1      | 1   | 14×14×384   |
| Pool3  | 3×3         | 2      | -   | 7×7×384     |
| FC1    | -           | -      | -   | 512         |
| FC2    | -           | -      | -   | 512         |
| Output | -           | -      | -   | 8           |

Figure 2. Images distribution of age groups.

4. Experiments
Our approach is implemented using Caffe [30], an open-source framework. Our implementations are based on AMD Ryzen 7 1800X Eight-Core 3.6GHz processor and Nvidia GeForce GTX 1080 Ti GPU.

4.1. Data preprocessing and Training
In order to decrease the influence of the background of images, the aligned images with 816×816 pixels are cropped to 256×256 pixels around the face center. The weights are initialized using random values from a zero mean Gaussian with standard deviation of 0.01. It must be noted that none of pre-trained models are utilized to initialize the network. Moreover, the network is trained without using
any additional data. We adopt stochastic gradient descendent (SGD) with a mini-batch size of 64 for training. The learning rate starts from 0.001, reduced to 0.0001 after 10000 iterations. Furthermore, the maximum number of iterations is 50000.

4.2. The Adience Dataset

We conduct experiments using the unconstrained Adience benchmark [15] for age classification. The Adience dataset consists of images automatically uploaded to Flickr from iPhone 5 or later smartphones without prior manual filtering. Compared with other unconstrained datasets, the Adience dataset is greater variable in terms of pose, lighting conditions and quality. The entire Adience contains over 26,000 images of 2,284 subjects.

In this study, we utilize the in-plane aligned version of Adience. Figure 2 shows the distribution of images from different age groups. And eight groups are defined with the age ranges as same as [5]. Training and testing are carried out using a standard 5-fold, subject-exclusive cross validation protocol, defined in [15]. Besides, experiments are implemented using the aligned and cropped faces, respectively.

4.3. Results

To evaluate the performance of the proposed approach in age classification, we utilize two different criterions, including exact accuracy and 1-off accuracy. Exact accuracy rate is the exact age-group classification result, while 1-off accuracy is off by one adjacent age-group i.e., the subject belongs to the group immediately older or immediately younger than the predicted group. To our best knowledge, facial features may change very little between oldest faces in one age class and the youngest faces of the subsequent class. We compare our approach against some state-of-the-art methods for age classification on the Adience dataset: 1) [15] proposed dropout-SVM to avoid overfitting for age estimation. 2) [5] adopted the same CNN model but it was based on the softmax loss function.

As can be seen from Table 2, the results listed are the mean accuracy ± standard error all age categories. The reported exact and 1-off accuracies in [15] are 45.1±2.6 and 79.5±1.4, respectively. The reported exact and 1-off accuracies in [5] are 50.7±5.1 and 84.7±2.2, respectively. These low accuracies of facial images demonstrate the difficulty involved with the Adience dataset. Finally, with the multi-class focal loss, our approach achieves the best results, not only improving the accuracy of aligned faces, but also improving the accuracy of cropped faces. It is evident that the performance of all models has been improved via preprocessing data.

Figure 3 provides a few examples of the age classification errors made by our approach on the unconstrained benchmark. It is clear that most of the mistakes are caused by extremely challenging viewing conditions of the Adience benchmark, such as low resolution, lighting conditions, and pose variations.

| Method                        | Exact (%) | 1-off (%)  |
|-------------------------------|-----------|------------|
| Best from[15]                 | 45.1 ± 2.6| 79.5 ± 1.4|
| Best from[5]                  | 50.7 ± 5.1| 84.7 ± 2.2|
| Our approach with aligned faces | 53.0 ± 6.5| 87.1 ± 2.7|
| Our approach with cropped faces | 54.0 ± 6.4| 88.2 ± 2.3|
Figure 3. Age misclassifications. The younger is mistakenly classified as the older in the top row, the older is mistakenly classified as the younger in the bottom row.

5. Conclusions
In this paper, we consider class imbalance as the primary issue on the Adience benchmark for age classification. To address this issue, we have proposed a new CNN model with the multi-class focal loss function to improve the performance of age classification, which applies a modulating term to the cross entropy loss to focus learning on minority examples. Extensive experiments on the Adience benchmark have demonstrated the effectiveness of the proposed approach. The absence of large scale training data could be an issue in age classification task. In future work, we plan to investigate the performance of more complex network architecture using more sufficient data.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (Grant Nos. 61602434 and 91646114), Youth Innovation Promotion Association CAS, No. 2017393, Talent Foundation of China West Normal University (No.463177), Chongqing key standard technologies innovation program of key industries under grant cstc2017zdcy-zdyfX0076 and Social Livelihood Foundation of Chongqing (No. cstc2017shmsA120010).

References
[1] Rothe R, Timofte R and Gool L V 2016 Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks *International Journal of Computer Vision* 1-14
[2] Geng X, Zhou Z H, Zhang Y, Li G and Dai H 2006 Learning from facial aging patterns for automatic age estimation. In: *ACM International Conference on Multimedia, Santa Barbara, Ca, USA, October*, pp 307-16
[3] Tistarelli M, Yadav D, Vatsa M and Singh R 2013 Short- and long-time ageing effects in face recognition: *IET Digital Library*)
[4] Ramanathan N, Chellappa R and Biswas S 2009 Age progression in Human Faces : A Survey *Journal of Visual Languages & Computing* 15
[5] Levi G and Hassnecer T 2015 Age and gender classification using convolutional neural networks. In: *Computer Vision and Pattern Recognition Workshops*, pp 34-42
[6] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks. In: *International Conference on Neural Information Processing Systems*, pp 1097-105
[7] Panis G and Lanitis A 2015 An Overview of Research Activities in Facial Age Estimation Using the FG-NET Aging Database *Journal of American History* 84 455-62
[8] Ricanek K and Tesafaye T 2006 MORPH: a longitudinal image database of normal adult age-progression. In: *International Conference on Automatic Face and Gesture Recognition*, pp 341-5
[9] Fu Y, Guo G and Huang T S 2010 Age Synthesis and Estimation via Faces: A Survey *IEEE Transactions on Pattern Analysis & Machine Intelligence* 32 1955-76
[10] Kwon Y H and Lobo N D V 1999 Age Classification from Facial Images. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pp 762-7
[11] Geng X, Zhou Z H and Smithmiles K 2007 Automatic age estimation based on facial aging patterns IEEE Transactions on Pattern Analysis & Machine Intelligence 29 2234-40
[12] Fu Y, Xu Y and Huang T S 2007 Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features. In: IEEE International Conference on Multimedia and Expo, pp 1383-6
[13] Cootes T F, Taylor C J, Cooper D H and Graham J 1995 Active Shape Models-Their Training and Application Computer Vision & Image Understanding 61 38-59
[14] Gunay A and Nabiyev V V 2008 Automatic age classification with LBP. In: International Symposium on Computer and Information Sciences, pp 1-4
[15] Eidinger E, Enbar R and Hassner T 2014 Age and Gender Estimation of Unfiltered Faces IEEE Transactions on Information Forensics & Security 9 2170-9
[16] Ylioinas M, Hadid A and Pietikäinen M 2012 Age Classification in Unconstrained Conditions Using LBP Variants. In: International Conference on Pattern Recognition, pp 1257-60
[17] Gunay A and Nabiyev V V 2007 Automatic Detection of Anthropometric Features from Facial Images. In: Signal Processing and Communications Applications, 2007. Siu 2007. IEEE, pp 1-4
[18] Guo G, Mu G, Fu Y and Huang T S 2009 Human age estimation using bio-inspired features. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 112-9
[19] Cao D, Lei Z, Zhang Z, Feng J and Li S Z 2012 Human Age Estimation Using Ranking SVM. In: Chinese Conference on Biometric Recognition, pp 324-31
[20] Guo G, Fu Y, Dyer C R and Huang T S 2008 Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression IEEE Transactions on Image Processing 17 1178-88
[21] Fu Y and Huang T S 2008 Human Age Estimation With Regression on Discriminative Aging Manifold IEEE Transactions on Multimedia 10 578-84
[22] Guo G and Mu G 2013 Joint estimation of age, gender and ethnicity: CCA vs. PLS. In: IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp 1-6
[23] Yi D, Lei Z and Li S Z 2014 Age Estimation by Multi-scale Convolutional Network. In: Asian Conference on Computer Vision, pp 144-58
[24] Ranjan R, Zhou S, Chen J C and Kumar A 2015 Unconstrained Age Estimation with Deep Convolutional Neural Networks. In: IEEE International Conference on Computer Vision Workshop, pp 351-9
[25] Zhang A K, Guo B L, Gao C C, Zhao D Z, Sun E M and Yuan F X 2017 Age group classification in the wild with deep RoR architecture. In: IEEE International Conference on Image Processing, pp 1587-91
[26] Chen S, Zhang C, Dong M, Le J and Rao M 2017 Using Ranking-CNN for Age Estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 742-51
[27] Lin T Y, Goyal P, Girshick R, He K and Dollar P 2017 Focal Loss for Dense Object Detection 2999-3007
[28] Aydogdu M F, Celik V and Demirci M F 2017 Comparison of Three Different CNN Architectures for Age Classification. In: IEEE International Conference on Semantic Computing, pp 372-7
[29] Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 Improving neural networks by preventing co-adaptation of feature detectors Computer Science 3 págs. 212-23
[30] Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, Karayev, Sergey, Long and Jonathan 2014 Caffe: Convolutional Architecture for Fast Feature Embedding 675-8