Age Estimation of Real-Time Faces Using Convolutional Neural Network

Olatunbosun Agbo-Ajala and Serestina Viriri(✉)

School of Mathematics, Statistics and Computer Sciences,
University of Kwazulu-Natal, Durban, South Africa
ajalabosun@gmail.com, viriris@ukzn.ac.za

Abstract. Age classification of an individual from an unconstrained real-time face image is rapidly gaining more popularity and this is because of its many possible applications from security control, surveillance monitoring to forensic art. Several solutions have been proposed in the past few years in solving this problem. Many of the existing traditional methods addressed age classification from face images taken from a controlled environment, only a few studied an unconstrained imaging conditions problem from real-time faces. However, deep learning methods have proven to be effective in solving this problem especially with the availability of both a large amount of data for training and high-end machines. In view of this, we propose a deep learning solution to age estimation from real-life faces. A novel six-layer deep convolutional neural network (CNN) architecture, learns the facial representations needed to estimate ages of individuals from face images taken from uncontrolled ideal environments. In order to further enhance the performance and reduce overfitting problem, we pre-trained our model on a large IMDB-WIKI dataset to conform to face image contents and then tuned the network on the training portions of MORPH-II and OIU-Adience datasets to pick-up the peculiarities and the distribution of the dataset. Our experiments demonstrate the effectiveness of our method for age estimation in-the-wild when evaluated on OIU-Adience benchmark that is known to contain images of faces acquired in ideal and unconstrained conditions, where it achieves better performance than other CNN methods. The proposed age classification method achieves new state-of-the-art results with an improvement of 8.6% (Exact) and 3.4% (One-off) acccuracy over the best-reported result on OIU-Adience dataset.

Keywords: Age estimation · Face images · Convolutional neural network · Deep learning

1 Introduction

Age estimation using face images is an interesting and a very challenging task [1,2]. The features from the face images are used to determine age, gender, ethnic background, and emotional state of people [3]. Among this set of features,

N. T. Nguyen et al. (Eds.): ICCCI 2019, LNAI 11683, pp. 316–327, 2019.
https://doi.org/10.1007/978-3-030-28377-3_26

age estimation can be particularly useful in many possible real-time applications [4] which include biometrics [3], security and surveillance [5], electronic customer relationship management [6], human-computer interaction [5], electronic vending machines [6], forensic art [7], entertainment [8], cosmetology [1] among others.

The conventional hand-crafted methods relied on the differences in dimensions of facial features [4], and face descriptors like local binary patterns and Gabor features [9,10]. Most of these techniques only designed classification methods for age estimation task that utilize face images captured under controlled conditions [11]; few of those methods are designed to handle the many challenges of unconstrained imaging conditions. The images in these categories have some variations which may affect the ability of the computer vision system to accurately estimate the age. More recently, convolutional neural network (CNN) based methods have proven to be effective for age estimation task due to its superior performance over existing methods. Availability of both large data for training and high-end machines, also help in the adoption of the deep CNN methods for age classification problems.

In this paper, we present an age estimation system (in Fig. 1) that uses a deep CNN method to estimate the age of face images of an individual taken from unconstrained real-time scenarios. The design is a six-layer network architecture of four convolutional layers and two fully connected layers pre-trained on a large IMDB-WIKI dataset and tuned on the training portions of MORPH-II and OIU-Adience datasets for further learning the traits of face images in each datasets. The proposed system includes three stages of image preprocessing phase that prepare the face images before being fed into the designed network for age classification process (details of image preprocessing method is presented in Sect. 3.1). The newly designed network was evaluated on OIU-Adience benchmark for age classification of unfiltered face images. The method outperforms the other methods in the literature, showing an improvement on the current state-of-the-art methods.

The remainder of this paper is structured as follows: Sect. 2 presents a review of the related works, Sect. 3 describes the proposed method, Sect. 4 presents the experiments while conclusion and future works are drawn in Sect. 5.

2 Related Works

The study of age estimation from face images has been in existence for decades. Various methods have been employed in the time past to address this problem, with varying levels of achievement. A comprehensive study of some of the past but recent approaches to age estimation from facial images are presented by Angulu *et al.* in [12].

Some of the very early past methods approached an age estimation problem by manually extracting the facial features using differences in facial features dimension. Although those methods have proven to be effective when classifying images from a constrained environment, only a few have attempted to address the problems that arise from real-time images with the variations in

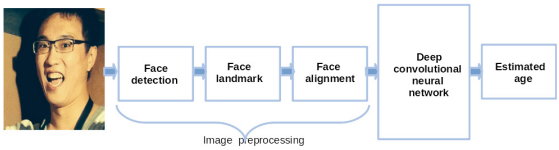

Fig. 1. A schematic diagram of the proposed age estimation system.

Fig. 2. Image preprocessing phase

pose, illumination, expression, and occlusion [4]. Kwon and Lobo [13] presented an age estimation solution that extracted and used wrinkles features. They used distance ratios between frontal face landmarks to separate babies from adults and separated the young adults from senior adults by using the wrinkle indices. However, their method lacks the ability to classify face images from in-the-wild scenarios due to the presence of varying degrees of variations in those images. Ramanathan and Chellappa in [14] proposed a model that predicted an age progression and face recognition of young faces from 0 to 18 years. They used images from FG-NET aging dataset and a private dataset for evaluation and testing. In [15] Horng *et al.* used both geometric and wrinkles features. They used geometric features to distinguish a baby face from an adult face and wrinkles for classifying adult faces into three different adult groups. They employed Sobel edge operator and region labeling to locate the positions of features to extract. The approach achieved a better result than the state-of-the-art methods on constrained face images. Jana *et al.* in [16] also investigated a method that used spatial local binary patterns (LBP) histograms to classify face images into six different age (groups). They employed minimum distance, nearest neighbor and k-nearest neighbour classifiers at the classification stage. The result showed a reasonable improvement on the existing age estimation methods on face images from controlled environments.

Although, all of these methods on age classifications have proven to be effective on constrained images, they are not suitable to tackle large variations experienced in an ideal real-world images. In order to effectively solve the task of age classification of real-time face images, increasing attention is drawn to the use of machine learning and deep CNN. Eidinger *et al.* in [4] developed a solution that estimates humans age from facial images acquired in a challenging in-the-wild condition. They collected face images labeled for age that are acquired from an

ideal world environment and employed a robust face alignment with a dropout-support vector machine approach to estimate the ages of individuals from face images taken from a real-time environment. Their approach significantly outperformed the state-of-the-art methods when evaluated. Levi and Hassner [11] investigated a deep CNN approach for age estimation from an unconstrained image. They developed a simple CNN architecture that can estimate ages of individuals using face images from real-time scenarios that reflect different levels of variations in appearance. Their method achieved 50.7% (Exact) and 84.7% (One-off) accuracy when it was evaluated on OIU-Adience dataset for age. Ekmekji [17] proposed a study that classified humans age by extending the already existing approaches. Qawaqneh *et al.* in [18] studied a solution that used an already trained deep CNN to estimate the age of unconstrained face images. The study used a network that was initially trained on face recognition dataset to carry out the age estimation task. Their approach outperformed the previous works when evaluated on the challenging OIU-Adience database. Liu *et al.* [19] also developed an approach that focused on the distribution of data rather than modifying the already existing network architectures. They proposed a CNN model that used a multi-class focal loss function instead of the conventional softmax function. Fortunately, their experimental approach showed better performance over the state-of-the-art techniques.

In summary, it has been proven in the literatures that CNN can achieve great success on age estimation task, significantly achieving better performance. Although most of the recent works improved classification accuracy by modifying the existing network architecture, age classification can still obtain a higher accuracy with a better CNN architecture. In this study, we propose a novel CNN method, a six-layer CNN architecture of four convolutional layers and two fully connected layers. The proposed method is fortified with a robust image processing technique that impact the performance of our approach. Furthermore, our optimization algorithm that adaptively tunes the learning rate as the network trains and a regularization method, contributed to the effectiveness of our approach. Our method resulting in a better performing network, showing an improvement on the state-of-the-art methods.

3 Proposed Method

The proposed method follows the pipeline in Fig. 1, in this section, we describe each step of the pipeline in detail.

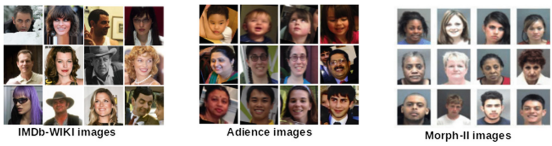

Fig. 3. Face images from IMDB-WIKI, Adience and MORPH-II datasets

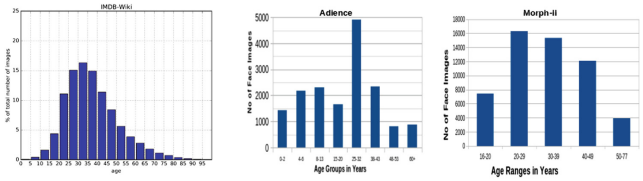

Fig. 4. Age distribution in IMDB-WIKI, Adience and MORPH-II datasets

3.1 Image Preprocessing

Some of the datasets employed in this work do not show centered frontal-faces but rather faces in-the-wild, we need to detect and align the faces for both training and testing. As such there is need to prepare and preprocess the face images for classification task before feeding them into the designed network. The image preprocessing phases are explained in more detailed below:

Face Detection: In order to detect the facial landmark of the input images, there is need to localize the face in the image and detect the key facial structures on the face. In this work, a face detector *Haar Feature-based Cascade Classifier* proposed by Viola and Jones [20] for face detection was employed. The classifier returns an output that is a bounding rectangle that contains the face image.

Landmark Detection: To represent salient facial regions like mouth, right eyebrow, left eyebrow, right eye, left eye, nose and jaw, a pre-trained Dlib model *shape_predictor_68_face_landmarks* that was an implementation of Kazemi and Sullivan [21] for face landmark detection was used. It estimates the location of $68(x, y)$-coordinates that map to facial structures on the face images. With this method, the key structures in our face images were localized.

Face Alignment: There is need to align the face images for both training and testing to further boost our work for higher accuracy. To this effect, there is need to compute the angle between the *(x, y)-coordinates* of the eyes, generated the midpoint between the eyes then applied affine transformation to warp the images into a new output coordinate space. With this, the face is centered in the image, rotated with the eyes lying along the same *y-coordinates* and then scaled with the size of all faces approximately equal. As expected, cropping the detected face for the age estimating processing rather than using the entire image, obtained a massive improvement in performance. Image preprocessing stage are shown in Fig. 2.

3.2 Network Architecture

We proposed an architecture (see Table 1) that contain six layers; four convolutional layers and fewer nodes with two fully-connected layers, The proposed network is a sequential CNN model that is capable of extracting the facial features needed to distinctively estimate the age of an individual via face image.

The method introduced a regularization technique (a dropout and data augmentation) to reduce the risk of overfitting and improve performance of our work. A batch normalization is used in place of the conventional local response normalization used by Levi *et al.* to further improve the performance of the network. We further prepare the face images by scaling them into 256 × 256 and then cropped into 227 × 227 pixel to boost the accuracy of our method. The six-layer CNN based architecture is structured as follow:

The first convolutional layer learned 96, 7 × 7 kernels with a stride of 4 × 4 to reduce the spatial dimensions of the input 227 × 227 images. Each convolutional layer is followed by an activation layer then a batch normalization with a max-pool of kernel size of 3 × 3 and a stride of 2 × 2 operating at the end of the convolutional block. After series of empirical experiments, a small dropout of 0.25 was utilized to reduce overfitting. The second series of convolutional layer applied the same structure, but with an adjustment to learn 256, 5 × 5 filters. The third is near identical to the other convolutional layers but with an increase in the number of filters to 384 and a reduction of the filter size to 3 × 3. The final convolutional layer set has a filter of 256 and a filter size of 3 × 3. All the convolutional layers are sandwiched with a dropout of 0.25 at the end of each layer set. The first fully-connected layer received the output of the fourth convolutional layer and learn 512 nodes with an activation layer followed by a batch normalization and with a dropout of 0.50 while the second and last fully connected layer maps to the final classes for age. In our case, we employed a softmax loss classifier to assign a probability for each class. The softmax classifier as a linear classifier, uses the cross-entropy loss function; the gradient of the cross-entropy function inform a softmax classifier how exactly to update its weights. The Softmax loss is defined in Eq. 1:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e_k^z} \tag{1}$$

where z is a vector of the inputs to the output layer and j indexes the output units.

For the cross-entropy loss of a multi-class classification, we calculate a separate loss for each class label per observation and sum the result. This is defined as presented in Eq. 2:

$$-\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{2}$$

where M is the number of classes, y is the binary indicator $(0or1)$ if class label c is the correct classification for observation o and p is the predicted probability observation o that is of class c.

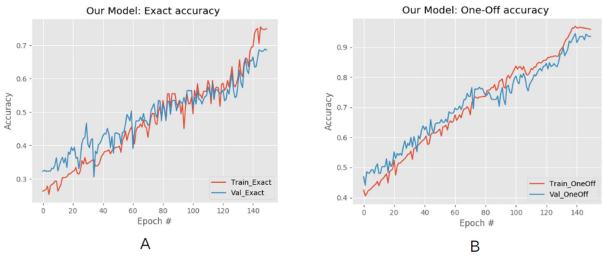

Fig. 5. Graph of Exact and One-off accuracy against Epoch

3.3 Evaluation Metrics

In order to evaluate our approach, we use two different established metrics.

Exact Accuracy: Exact accuracy metric is used to define the effectiveness of an age estimator. It is calculated as the percentage of face images that were classified into correct age-groups. Equation 3 presents its mathematical equation.

$$Exact\ accuracy = \frac{no\ of\ accurate\ prediction}{total\ no\ of\ prediction\ made} \tag{3}$$

One-Off: One-off evaluation metric measures whether the ground-truth class label matches the predicted class label. It allows for a deviation of at most one bucket from the real age range. One-off is calculated as a ratio of the correct predictions to the total number of data points.

3.4 Network Training

The proposed network is trained on IMDB-WIKI dataset, a very challenging dataset, to conform to face image contents and then tuned on the training portion of both MORPH-II and OIU-Adience datasets to pick-up the peculiarities and the distribution of each dataset. All these datasets are with varying degrees of variations, as such, there is need to preprocess the face images to decrease the influence of the background of images and resized the images to 256×256 pixel. Data augmentation is also needed to increase the amount of relevant data within a training dataset and reduce the risk of overfitting of the network, and this is done by creating an altered copies of the face images during the training stage. For the estimation, we utilized 70% OIU-Adience for training, and the other 30% for validation and testing. To further boost the accuracy of our approach, 10-crop oversampling method was employed to extract 227×227 region of the images and this increased the training data and consequently improved our result. Moreover, rather than computing the loss and the gradient of the entire training set, a sample of small sets of training examples (mini batch) was used to compute the estimate of the full sum and that of the true gradient, a Stochastic Gradient Descendent (SGD) optimizer was adopted with a mini-batch size of 64 for training. The optimizer was chosen ahead of other known ones in order

Table 1. Summary of our Network Architecture

Layer type	Output size	Filter size/Stride
INPUT IMAGE	$227 \times 227 \times 3$	–
CONV1	$56 \times 56 \times 96$	$7 \times 7/4 \times 4$
ACT	$56 \times 56 \times 96$	-
BN	$56 \times 56 \times 96$	-
Maxpool	$28 \times 28 \times 96$	$3 \times 3/2 \times 2$
dropout	$28 \times 28 \times 96$	-
CONV2	$28 \times 28 \times 256$	5×5
ACT	$28 \times 28 \times 256$	-
BN	$28 \times 28 \times 256$	-
Maxpool	$14 \times 14 \times 256$	3×3
dropout	$14 \times 14 \times 256$	-
CONV3	$14 \times 14 \times 384$	3×3
ACT	$14 \times 14 \times 384$	-
BN	$14 \times 14 \times 384$	-
Maxpool	$7 \times 7 \times 384$	3×3
dropout	$7 \times 7 \times 384$	-
CONV4	$7 \times 7 \times 384$	3×3
ACT	$7 \times 7 \times 256$	-
BN	$7 \times 7 \times 256$	-
Maxpool	$1 \times 1 \times 256$	3×3
dropout	$1 \times 1 \times 256$	-
FC1	512	-
ACT	512	-
BN	512	-
dropout	512	-
FC2	8	-

to compute an update for each example $(x^{(i)}, y^{(i)})$ that was uniformly sampled from the training dataset; it calculates the gradient of the parameters by using only a few training samples. This is calculated as shown in Eq. 4 below:

$$\theta = \theta - \alpha \nabla_\theta J(\theta; x^{(i)}, y^{(i)}) \tag{4}$$

where α is the learning rate, $\nabla_\theta J$ is the gradient of the loss term with respect to the weight vector θ.

Data Augmentation: Figure 4 reveals the uneven distribution of the age in the employed datasets. To address this problem, we employed an adaptive augmentation method that increases the number of altered copies of the face images and

also makes the age distribution of the training set even. In this work, we applied an augmentation approach that includes random cropping, zooming, random mirror, and rotation.

Table 2. Age classification: Exact and One-off results on OIU-Adience benchmark.

Methods	Exact (%)	One-off (%)
Eidinger *et al.* [4]	45.1	79.5
Ekmekji [17]	54.5	84.1
Levi *et al.* [11]	50.7	84.7
Liu *et al.* [19]	54.0	88.2
Proposed	**63.1**	**91.6**

4 Experiments

In this section, we present the results of our empirical experiments. We introduce the datatsets used and then show the performance of our proposed method on the validation datasets.

4.1 Datasets

The availability of relevant facial aging databases plays an important role in the performance of an age estimator. In this work, we employed the three most relevant facial aging databases to either train, or validate our approach.

IMDB-WIKI: IMDB-WIKI [22] is the largest publicly-available dataset for age estimation of people in-the-wild, containing more than half a million images with accurate age labels between 0 and 100 years. IMDB contains 460,723 images of 20,284 celebrities and Wikipedia with 62,328 images. The images of IMDB-WIKI dataset are obtained directly from the website, as such the dataset contains many low-quality images, such as human comic images, sketch images, severe facial mask, full body images, multi-person images, blank images, and so on.

MORPH-II: MORPH-II database [23] is a publicly-available aging database collected at the University of North Carolina at Wilmington by the face aging group. The database records linked attributes such as age, gender, ethnicity, weight, height, and ancestry. The whole database is divided into two albums; album I and album II. Album II contains 55,134 face images obtained from more than 13,000 subjects.

OIU-Adience: OIU-Adience database [4] is a collection of face images from an ideal real-life and unconstrained environments. It reflects all the features that are expected of an image collected from challenging real-world scenarios. OIU-Adience images, therefore, exhibit a high level of variations in noise, pose,

appearance among others. It is used in studying age and gender classification system. The entire collection of OIU-Adience database is about 26,580 face images of 2,284 subjects and with an age-group label of eight comprising: 0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, 60+. Samples of the face images are presented in Fig. 3.

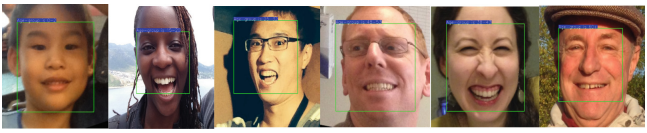

Fig. 6. Samples of face images with correct estimation

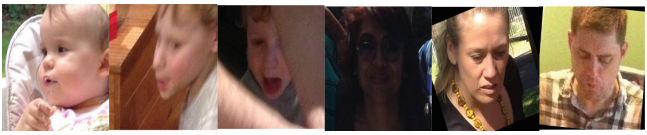

Fig. 7. Samples of face images with wrong estimation

4.2 Experimental Result

In this section, we assess our method for predicting age-groups. The purpose is to predict whether a person's age falls within some age range rather than predicting the precise age. We evaluate the performance of our classifier on OIU-Adience dataset using Exact and One-off accuracy metrics. We achieved an Exact accuracy of 63.1% and One-off accuracy of 91.6%. The graph in Fig. 5 presents the results for Exact accuracy and One-off accuracy. Table 2 presents our result and the results of the current state-of-the-art methods for age-group classification on OIU-Adience benchmark. Our approach achieves the best results, not only improving the Exact accuracy but also the One-off accuracy; it outperforms the current state-of-the-arts methods.

In Figs. 6 and 7, we present the predictions of some of the face images from the OIU-Adience (validation set) by our classifier. In many instances, our solution is able to correctly predict the age-group of faces. Failures (Fig. 7) may be as a result of two major reasons: The first is the failure to either detect or align the face. The second is because of some extreme conditions of variability such as non-frontal, blurring, low resolution, occlusion, heavy makeup.

5 Conclusions and Future Work

The proposed six-layer CNN based method shows the state-of-the-art result on OIU-Adience dataset. With a robust image processing design, our method

handles some of the variability noticed in the face images from real life's scenarios. This validates the applicability of our method to age classification in-the-wild. Pre-training on the IMDB-WIKI and fine-tuning on MORPH-II and OIU-Adience datasets, result in a large boost in the performances. In the future, we hope that a larger dataset will be available in age (groups) estimation so as to employ a deeper convolutional neural network architecture for age estimation of real-time faces. Fine-tuning the face detector on the target dataset(s) can also lessen the failure rate of the face detection phase. A more robust landmark detector can also lead to better alignment and performance.

References

1. Drobnyh, K.A., Polovinkin, A.N.: Using supervised deep learning for human age estimation problem. In: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W4, pp. 97–100, May 2017
2. Huerta, I., Fernández, C., Segura, C., Hernando, J., Prati, A.: A deep analysis on age estimation. Pattern Recogn. Lett. **68**, 239–249 (2015)
3. Bouchrika, I., Harrati, N., Ladjailia, A., Khedairia, S.: Age estimation from facial images based on hierarchical feature selection. In: 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering - STA 2015, Monastir, Tunisia, pp. 393–397 (2015)
4. Eidinger, E., Enbar, R., Hassner, T.: Age and gender estimation of unfiltered faces. IEEE Trans. Inform. Forensics Secur. **9**(12), 2170–2179 (2014)
5. Abbas, A.R., Kareem, A.R.: Intelligent age estimation from facial images using machine learning techniques. Iraqi J. Sci. **59**(2A), 724–732 (2018)
6. Mandal, S., Debnath, C., Kumari, L.: Automated age prediction using wrinkles features of facial images and neural network. Int. J. Emerg. Eng. Res. Technol. **5**(2), 12–20 (2017)
7. Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., Yuille, A.: Deep regression forests for age estimation. In: CVPR, pp. 2304–2313 (2017)
8. Wen, Y., Liu, W., Yang, M., Fu, Y., Xiang, Y., Hu, R.: Structured occlusion coding for robust face recognition. Neurocomputing **178**, 11–24 (2016)
9. Badame, V., Jamadagni, M.: Study of approaches for human facial age. Int. J. Innov. Res. Sci. Eng. Technol. **6**(8), 2347–6710 (2017)
10. Feng, S., Lang, C., Feng, J., Wang, T., Luo, J.: Human facial age estimation by cost-sensitive label ranking and trace norm regularization. IEEE Trans. Multimedia **19**(1), 136–148 (2017)
11. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–42 (2015)
12. Angulu, R., Tapamo, J.R., Adewumi, A.O.: Age estimation via face images: a survey. EURASIP J. Image Video Process. **2018**, 42 (2018)
13. Kwon, Y.H.: Age classification from facial images. Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki **74**(1), 1–21 (1997)
14. Ramanathan, N., Chellappa, R.: Modeling age progression in young faces. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006, February 2016

15. Horng, W.B., Lee, C.P., Chen, C.W.: Classification of age groups based on facial features. Tamkang J. Sci. Eng. **4**(3), 183–192 (2001)
16. Jana, R., Pal, H., Chowdhury, A.: Age group estimation using face angle. IOSR J. Org. **7**(5), 1–5 (2012)
17. Ekmekji, A.S.U.: Convolutional Neural Networks for Age and Gender Classification Research paper (2016)
18. Qawaqneh, Z., Mallouh, A.A., Barkana, B.D.: Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model. arXiv, September 2017
19. Liu, W., Chen, L., Chen, Y.: Age classification using convolutional neural networks with the multi-class focal loss. In: IOP Conference Series: Materials Science and Engineering, vol. 428, no. 1 (2018)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol. 1 (2001)
21. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
22. Zhang, K., et al.: Age group and gender estimation in the wild with deep RoR architecture. IEEE Access **5**(X), 22492–22503 (2017)
23. Ricanek, K., Tesafaye, T.: MORPH: a longitudinal image database of normal adult age-progression. In: Proceedings of 7th International Conference on Automatic Face and Gesture Recognition, pp. 341–345 (2006)