Enhanced Convolutional Neural Network for Age Estimation

Idowu Aruleba and Serestina Viriri[✉]

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
`viriris@ukzn.ac.za`

Abstract. The human face constitute various biometric features that could be used to estimate an important detail such as age. Variations in facial landmarks and appearances have presented challenges to automated age estimation. This account for limitations attributed to conventional approaches such as the traditional hand-crafted method, which cannot efficiently and adequately estimate age. In this study, a six layered Convolutional Neural Network (CNN) were proposed, which extract features from facial images taken in an uncontrolled environment, and classifies them into appropriate classes. Since a huge datasets is needed to obtain good accuracy from the trained model and minimize overfitting, data augmentation was performed on the datasets to balance the number of images in each class. The UTKFace dataset was used to train the model while validation was carried out on FGNET dataset. With the proposed novel method, an accuracy of 89.75% was recorded on the UTKFace dataset, which is a significant improvement over existing state-of-the-art methods previously implemented on the UTKFace dataset.

Keywords: Age estimation · Pre-processing · Feature extraction · Convolutional Neural Network

1 Introduction

The use of biometric features to determine a person's age is described as age estimation and can be accomplished using traits (biometric) such as eye color, skin, and facial emotions [1]. In the last few years, numerous research has been conducted in age estimation using facial images for real-world applications such as access monitoring and control, enhanced communication. These impacts have contributed significantly to emotiona1 intelligence, security control, and entertainment, among several others [1–3].

Many studies have been carried out to implement age estimations effectively. Although there still exist several limitations and drawbacks [6]. Previous approaches have real-world limitations majorly due to variations in conditions in which some images are taken, and could, in turn, affect the accuracy of the

estimations. These conditions include lighting, background, and facial expression, which, cumulatively, could account for reductions in the quality of images to be processed [4,5].

Several techniques have been previously employed, some of which include hand-craft methods for extracting visual features [5,6], different age modeling approaches [7] for image representation, and machine-based approaches for resolving problems with classification [14] and regression [15]. Most of these approaches have one or two weaknesses, like considering the distance between facial landmarks and neglecting the facial appearance, not able to encode wrinkles on the face, in addition to their inability to match boundaries in an image [6].

In this work, we propose an efficient deep learning approach that would serve as an end-to-end model to increase the accuracy of deep neural networks in the age estimation task. Convolutional Neural Network is known for its ability to learn from raw pixels, exploit spatial correlations in data, among several functionalities with proven high performance in the area of computer vision and image processing. The proposed model will be trained using a large benchmark dataset with unconstrained faces, while the accuracy will be compared to other state-of-the-art approaches.

This paper is structured as follows: in Sect. 2, we described the background and related work; in Sect. 3, we describe the proposed Convolutional Neural Networks method and its application in age estimation; Sect. 4, we provide detailed analyses of the experimental result, and ultimately, the conclusions are presented in Sect. 5.

2   Background and Related Work

The possibility of performing age estimation using facial images based on biometric features extracted from a person's face has been previously established. A survey carried out by Angulu, R. et al. [6] provided extensive details based on recent researches in aging and age estimation, which implemented popular algorithms of existing models with comparative analyses of their performance.

Deep learning methods like the deep convolutional neural network has performed excellently on different tasks of computer vision and image recognition such as age & gender classification [9], diagnosing Tuberculosis [12] because of the strength to learn without explicitly working on and ability to deal with extensive data [4,13]. In 2016, Zhang, Z. [2] study was based on apparent age estimation with CNN using Google-NET architecture [20]. A batch normalization layer was added to the model after each rectified linear unit operation. However, their method did not work well on older people, blurry images, and misaligned faces.

Anand, A. et al. [1], previously modeled a CNN method to cumulatively perform feature level fusion, dimension reduction of feature space, and individual age estimation. WIKI dataset [21] was employed to model the Feed Forward Neural Networks (FFNNs) for parameter estimation prior to training. Adience Benchmark dataset was used to compare the exact accuracy with the state-of-the-art techniques.

Mualla, N. et al. [10], in their study, made use of PCA for feature extraction and reduction. Thereafter, the Deep Belief Network (DBN) classifier was compared with the K-nearest neighbor algorithm and Support Vector Machine classifiers. From their results, the DBN accuracy for age estimation output significantly outperforms the KNN and SVM classifier.

Niu, Z. et al. [16] presented a multiple output CNN learning algorithm with ordinal regression to solve the classification problem for age estimation [16]. However, Yi, D. et al. [17] proposed the use of the multi-scale convolutional network for age estimation. In their study, they proposed a CNN model which was not vast enough. It consisted of only one fully connected layer, one convolution layer, one pooling, and one local layer, which summed up to a total of four layers, while just a subset of MORPH2 was used to train the model.

Sendik, O. et al. [8] applied Convolutional Neural Network (CNN) and Support Vector Regressor (SVR) in their face-based age estimation task. They trained the CNN for representation and metric learning, the SVR was applied to the learned features. The work demonstrated that by retraining the SVR layer, small datasets can be analysed.

Huang, J. et al. [11] proposed a deep learning technique for age group classification using the Deep Identification features, the architecture was optimized for face representation. They analysed their architecture on both constrained and unconstrained databases which performs well.

The study carried out by Das, A. et al. [19] analyzes the classification of gender, age, and race as a multi-task. They proposed a multi-task Convolutional Neural Network (MTCNN), which makes use of unconnected features of the fully connected layers. They split the fully connected layers to perform multi-task learning for a better face attribute analysis efficiently. However, they took advantage of the hierarchical distribution of information that is present in CNN features. The experiment was carried out using two datasets; UTKFace and Bias Estimation in Face Analytics (BEFA). Their approach provided a promising outcome on both datasets.

3   Methods and Techniques

3.1   Datasets

The following databases were used for the experiments done in this work. These databases are publicly available at:

UTKFace
The UTKFace dataset is a large face dataset that consists of over 20,000 face images in the wild with a long span ranging from 0 to 116 years old. Each facial image is with annotations of age, gender, and ethnicity. Some face samples provide corresponding aligned and cropped faces. The images cover large variations in pose, facial expression, illumination, occlusion, and resolution, to mention but

a few. The UTKFace dataset has been used for different tasks e.g. age estimation [19], age progression/regression [22] etc. In this study, the ages of the face images are grouped, as shown in Table 1.

The dataset can be accessed at https://www.kaggle.com/jangedoo/utkface-new.

Table 1. Ages from the dataset as grouped

Age group	Age range	Number of images
Group 1	0–18	3943
Group 2	19–30	7377
Group 3	31–80	9477
Group 4	80 above	491

FGNET
FGNET dataset only contains 1002 images of 82 individuals in the wild; the ages in dataset ranges from 0 to 69 years old with about 12 images per person. This is a small dataset with every constituent images representing different ages.

3.2 PreProcessing

Working with a deep neural network requires a large data size to achieve higher accuracy in the estimation of age. As such, we carried out data augmentation on our dataset to balance the number of images for each age class. In Table 1, age class ≥ 80 compared to age class 31–80 shows the differences in the number of images. These differences could adversely affect the accuracy of our model. For the augmentation process, we applied different augmentations like flip left right (0.5), skew $(0.4, 0.5)$ and zoom (probability $= 0.2$, min factor $= 1.1$, max factor $= 1.5$).

3.3 Proposed Model

Convolutional Neural networks are deep neural networks for classification tasks. The novel CNN architecture in this work consists of four convolutional layers and two fully connected layers. The input image with shape $200 \times 200 \times 1$, which was passed in the first convolutional layer learning at 32, Asides from the second convolutional layer learning 64, which has a 5×5 filter, other convolution layers consist of a 3×3 filter, the same padding, and activation function (rectified linear unit). The four convolutional layers made use of Max pooling with 2 * 2 poling size with a dropout 0.25 and also a batch normalization layer. The first fully connected layer consists of 512 neurons, a Rectified Linear Unit activation, then a dropout ratio of 0.5. The second fully connected layer has four (4) neurons with a Softmax as the activation function. After features have been extracted from the input images at the convolutional layer, a softmax classifier is needed to determine the probability of the feature extracted and return the confidence
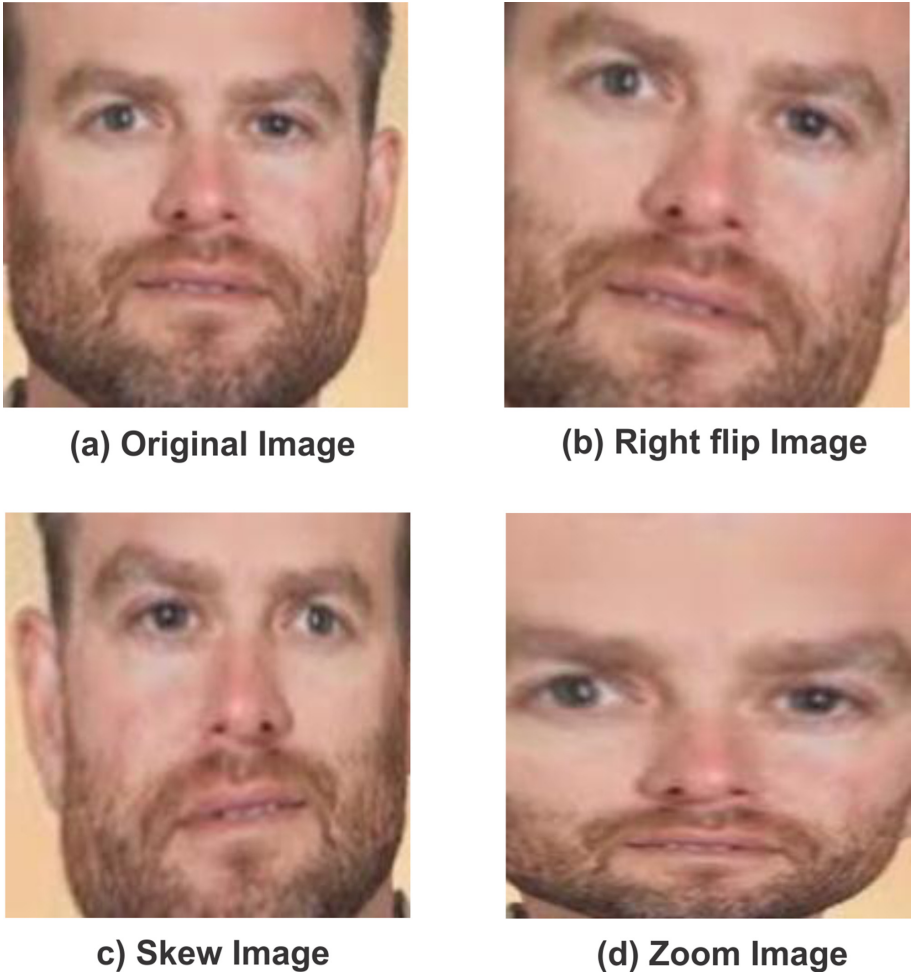
Fig. 1. Sample of the preprocessed images for the flip, skew, and zoom.
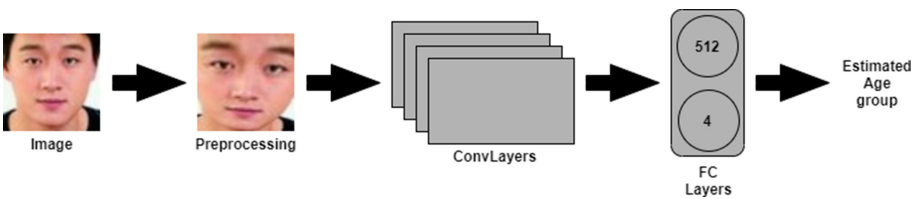

Fig. 2. The architecture of the proposed age estimation.

score for each class. The mathematical representation of the softmax function is as follows (Figs. 1 and 2);

$$S(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{k} e_j^x}$$

where each value of x in the input vector is an exponent divided by a sum of the exponents of all the inputs.

Our model was trained using Adam optimizer with an initial learning rate of 0.0001, and we used the reduced learning rate on Plateau to monitor the validation accuracy and loss setting the patience at ten (10) and the minimum learning rate at 0.001. The Adam optimizer is given as below;

$$\theta_{t+1} = \theta_t - \frac{\eta \cdot \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

and where

$$m_t = (1 - \beta_1)g_t + \beta_1 m_{t-1}, v_t = (1 - \beta_2)g_t^2 + \beta_2 v_{t-1}$$

where ϵ is Epsilon, η is the learning rate, g is the gradient (Table 2).

Table 2. The training details with UTKFace dataset

Classifier	Optimizer	No of epoch	Initial learning rate	Weight decay
Age group	Adam	35	0.0001	0.005

During the implementation, fewer convolutional filters should be used in the initial layer and increased filter in subsequent layers. Furthermore, wider window size in the initial layers compared to the later layers. Finally, the use of dropout can help to avoid over-fitting (Table 3).

4   Result and Discussion

The overall performance of the proposed model is evaluated using the accuracy metric, and the system achieves an accuracy rate of 89.75% on the classifier. Figure 3 shows the accuracy achieved with this method. It can be seen that the model accuracy curve goes optimal before five epochs. The categorical cross-entropy used for the evaluation of the loss function in the classifier achieves a loss of 0.3389 on the validation dataset, as shown in Fig. 3. Our model is adept to correctly estimate the age group of faces in many instances, while in some instances of estimating the age group, misclassification was recorded.

Table 3. The details of the architecture for the proposed method

Layers	Output size	Filter size
Input image	$200 \times 200 \times 1$	
CONV1	$200 \times 200 \times 32$	3×3
BATCHNORM	$200 \times 200 \times 32$	
ACT	$200 \times 200 \times 32$	
MAXPOOLING	$100 \times 100 \times 32$	
DROPOUT	$100 \times 100 \times 32$	
CONV2	$100 \times 100 \times 64$	5×5
BATCHNORM	$100 \times 100 \times 64$	
ACT	$100 \times 100 \times 64$	
MAXPOOLING	$50 \times 50 \times 64$	
DROPOUT	$50 \times 50 \times 64$	
CONV3	$50 \times 50 \times 128$	3×3
BATCHNORM	$50 \times 50 \times 128$	
ACT	$50 \times 50 \times 128$	
MAXPOOLING	$25 \times 25 \times 128$	
DROPOUT	$25 \times 25 \times 128$	
CONV4	$25 \times 25 \times 256$	3×3
BATCHNORM	$25 \times 25 \times 256$	
ACT	$25 \times 25 \times 256$	
MAXPOOLING	$12 \times 12 \times 256$	
DROPOUT	$12 \times 12 \times 256$	
FULLY CONN1	512	
BATCHNORM	512	
ACTIVATION	512	
DROPOUT	512	
FULLY CONN2	4	

Relatively, our model demonstrated a higher accuracy when implemented on the UTKFace dataset typifying the state-of-the-art method when compared to the accuracy reported in the study by Sithungu et al. [23] and Das et al. [19] (see Table 4). Sithungu et al. [110], in their study, employed a lightweight CNN model for age estimation analyses, after which the UTKFace dataset was used to train and test the performance of their model. Only 7700 out of 23718 images were used, which presents a limitation to their study since training a model with few data causes overfitting. Contrary to ours, we achieved augmentation of data in our study, which enabled us to use a large dataset to train and test our classifier. Das et al. [19] proposed the Multi-Task Convolution Neural Network
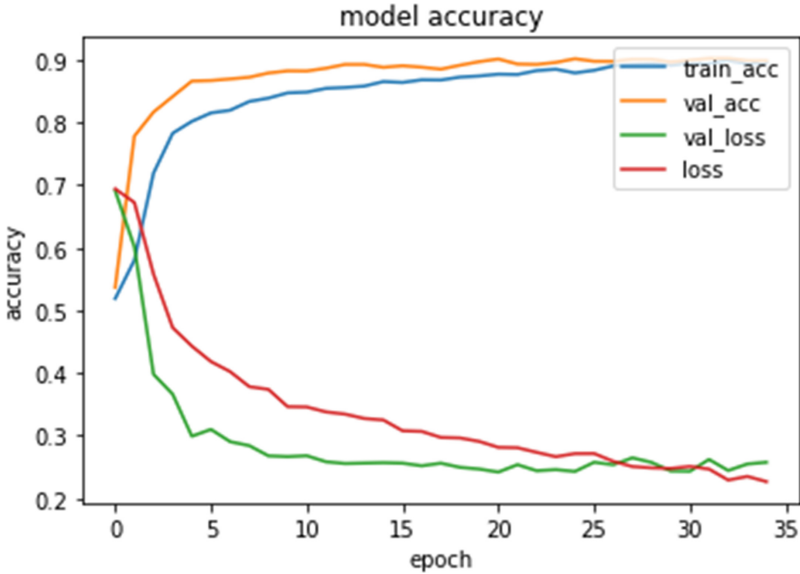
Fig. 3. Model accuracy and loss.

Table 4. Proposed model as compared

REF	Accuracy (%)
[23]	*45.3*
[18]	*63.97*
[19]	*70.1*
Proposed model	*89.75*

(MTCNN) using joint dynamic weight loss alteration to classify gender, age, and race jointly. While they achieved high gender and race precision, their age estimate was less accurate than our model.

5   Conclusion

The major challenge associated with age classification has been the stochastic nature of aging among individual and uncontrollable age progression information displayed on faces as deduced from models applied in previous studies. This validates the need for a more efficient model that can accurately extract features from input images, which is the major underlying process in image classification. Though our CNN approach is a custom one, but with the introduction and application of a large facial image dataset we consider that our model can achieve a meaningful improvement. In this study, the proposed model is implemented on the UTKFace dataset for extracting distinctive features and then classified into

the appropriate classes. The efficiency of the model is being validated on the FGNET dataset, and the performance is compared with other studies, as shown in Table 4. The proposed CNN architecture performed very well and demonstrated the state-of-the-art result on the UTKFace dataset. For future works, we would consider the effect of gender information on age estimation, consider a deeper CNN model age, gender and ethnicity. Also, we would explore some pretrained deep learning model to our study.

References

1. Anand, A., Labati, R.D., Genovese, A., Muñoz, E., Piuri, V., Scotti, F.: Age estimation based on face images and pre-trained convolutional neural networks. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–7. IEEE (2017)
2. Zhang, Z.: Apparent, age estimation with CNN. In: 4th International Conference on Machinery, Materials and Information Technology Applications, p. 2017. Atlantis Press (2016)
3. Agbo-Ajala, O., Viriri, S.: Age estimation of real-time faces using convolutional neural network. In: Nguyen, N.T., Chbeir, R., Exposito, E., Aniorté, P., Trawiński, B. (eds.) ICCCI 2019. LNCS (LNAI), vol. 11683, pp. 316–327. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28377-3_26
4. Liu, H., Lu, J., Feng, J., Zhou, J.: Ordinal deep learning for facial age estimation. IEEE Trans. Circ. Syst. Video Technol. **29**(2), 486–501 (2017)
5. Huerta, I., Fernández, C., Segura, C., Hernando, J., Prati, A.: A deep analysis on age estimation. Pattern Recogn. Lett. **68**, 239–249 (2015)
6. Angulu, R., Tapamo, J.R., Adewumi, A.O.: Age estimation via face images: a survey. EURASIP J. Image Video Process. **2018**(1), 42 (2018)
7. Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. Comput. Vis. Image Underst. **74**(1), 1–21 (1999)
8. Sendik, O., Keller, Y.: DeepAge: deep learning of face-based age estimation. Sig. Process. Image Commun. **78**, 368–375 (2019)
9. Agbo-Ajala, O., Viriri, S.: Deeply learned classifiers for age and gender predictions of unfiltered faces. Sci. World J. **2020**, 1289408 (2020)
10. Mualla, N., Houssein, E.H., Zayed, H.H.: Face age estimation approach based on deep learning and principle component analysis. Int. J. Adv. Comput. Sci. Appl. **9**(2), 152–157 (2018)
11. Huang, J., Li, B., Zhu, J., Chen, J.: Age classification with deep learning face representation. Multimedia Tools Appl. **76**(19), 20231–20247 (2017)
12. Oloko-Oba, M., Viriri, S.: Diagnosing tuberculosis using deep convolutional neural network. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds.) ICISP 2020. LNCS, vol. 12119, pp. 151–161. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51935-3_16
13. Qiu, J.: Convolutional neural network based age estimation from facial image and depth prediction from single image (2016)
14. Geng, X., Yin, C., Zhou, Z.H.: Facial age estimation by learning from label distributions. IEEE Trans. Pattern Anal. Mach. Intell. **35**(10), 2401–2412 (2013)
15. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. **24**(4), 442–455 (2002)

16. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output CNN for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4920–4928 (2016)
17. Yi, D., Lei, Z., Li, S.Z.: Age estimation by multi-scale convolutional network. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9005, pp. 144–158. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16811-1_10
18. Savchenko, A.V.: Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet. PeerJ Comput. Sci. **5**, e197 (2019)
19. Das, A., Dantcheva, A., Bremond, F.: Mitigating bias in gender, age and ethnicity classification: a multi-task convolution neural network approach. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 4920–4928 (2018)
20. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
21. Rothe, R., Timofte, R., Van Gool, L.: Deep expectation of real and apparent age from a single image without facial landmarks. Int. J. Comput. Vis. **126**(2), 144–157 (2016). https://doi.org/10.1007/s11263-016-0940-3
22. Zeng, J., Ma, X., Zhou, K.: CAAE++: improved CAAE for age progression/regression. IEEE Access **6**, 66715–66722 (2018)
23. Sithungu, S., Van der Haar, D.: Real-time age detection using a convolutional neural network. In: Abramowicz, W., Corchuelo, R. (eds.) BIS 2019. LNBIP, vol. 354, pp. 245–256. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20482-2_20