

A Lightweight Convolutional Neural Network for Real and Apparent Age Estimation in Unconstrained Face Images

OLATUNBOSUN AGBO-AJALA¹ AND SERESTINA VIRIRI¹, (Member, IEEE)

School of Mathematics, Statistics, and Computer Science, University of Kwazulu-Natal, Durban 4000, South Africa

Corresponding author: Serestina Viriri (viriris@ukzn.ac.za)

ABSTRACT Real and apparent age estimation of human face has attracted increased attention due to its numerous real-world applications. Different intelligent application scenarios can benefit from these computer-based systems that predict the ages of people correctly. Automatic apparent age system is particularly useful in medical diagnosis, facial beauty product development, movie role casting, the effect of plastic surgery, and anti-aging treatment. Predicting the real and apparent age of people has been quite difficult for both machines and humans. More recently, Deep learning with Convolutional Neural Networks (CNNs) methods have been extensively used for these classification task. It has incomparable advantages in extracting discriminative image features from human faces. However, many of the existing CNN-based methods are designed to be deeper and larger with more complex layers that makes it challenging to deploy on mobile devices with resource-constrained features. Therefore, we design a lightweight CNN model of fewer layers to estimate the real and apparent age of individuals from unconstrained real-time face images that can be deployed on mobile devices. The experimental results, when analyzed for classification accuracy on FG-NET, MORPH-II and APPA-REAL, with large-scale face images containing both real and apparent age annotations, show that our model obtains a state-of-the-art performance in both real and apparent age classification when compared to state-of-the-art methods. The new results and model size, therefore, confirm the usefulness of the model on resource-constrained mobile devices.

INDEX TERMS Apparent age, unconstrained images, convolutional neural network, mobile devices.

I. INTRODUCTION

Age estimation of faces is a very prolific area of research within the computer vision community [1], [2]. There has been an increasing interest toward age estimation from facial images [1] due to its increasing demands in various potential applications including security control [3], human-computer interaction [3], social media [4] and forensic studies [5], [6]. Although this subject has been extensively studied, the ability to estimate human ages reliably and correctly from face images is still far from satisfying human performance levels [7]. There exist two kinds of facial age estimation. One is real (biological) age estimation, which determines the precise chronological or biological age of a person from the facial image [8]; the other is apparent age estimation [9], this focuses on “how old does a person looks like” rather than

predicting the real or biological age. The difference between the traditional real age estimation and apparent age estimation is that the age labels in apparent are annotated by human assessors rather than the real biological age. In reality, some people may appear younger than their real age, while some may appear older. As a result, the real age may be quite different from the apparent age for each subject.

Several methods have been proposed for these task. Recently Convolutional Neural Networks (CNNs) has shown some satisfactory performance in different areas of face-related analysis, including face alignment [10], face recognition [11]–[13], face verification [14], age and gender classification [15], [16]. Deep learning and CNN have been effectively employed to transform the face images to output the age label [17]. This approach is different from handcrafted-based methods that manually extract the features from the face. In CNN, the discriminative features are automatically extracted during the training procedure rather than

The associate editor coordinating the review of this manuscript and approving it for publication was Hossein Rahmani¹.

defining a set of algorithms. It solves the estimation problem in an hierarchical manner of concept by encoding the basic representation in the lower layers, and form the more abstract concepts in the higher layers using those concepts received from the low layers [18].

However, many of the state-of-the-art CNN models are larger and more complex, with too large network parameters and layers, very long training time and huge training dataset, which bring expensive computation cost and storage overhead. Hence, it ultimately makes it challenging to deploy those models on mobile devices with resource-constrained features. Therefore, we design a multi-task lightweight CNN model that accurately estimates the real and apparent age of human faces, which can also be deployed on mobile terminals.

The contributions of this paper are summarized as follows:

- 1) We propose a multi-task lightweight convolutional neural network model for real and apparent age estimation of unconstrained faces.
- 2) We develop an age estimation model that is easy and effective and can be deployed on resource-constrained mobile devices.
- 3) We also design an image preprocessing method that includes face and landmark detection on the unconstrained input face images to prepare the image.
- 4) The model comprises the image augmentation (regularization) process that produces alter copies of the images, which increase the feature information related to the face age.

The remainder of the paper is organized as follows. Section 2 briefly reviews the previous but closely related works in real and apparent age estimation. Section 3 describes the proposed approach in detail. Section 4 provides the experimental results of our method and discussion. Finally, Section 5 concludes the paper and presents the future works.

II. RELATED WORKS

In this section, we present a review of the related works in real and apparent age estimation.

A. REAL AGE

Many great models have been developed in recent years for real age estimation tasks. For example, in 2015, Levi and Hassner [15] developed a shallow CNN architecture that used “three convolutional layers and two fully connected layers” to learn features representations. Agustsson *et al.* [9] proposed a “deep Residual Deep Expectation” (DEX) method that possess the capacity to improve the performance of the original “DEX regressors” on age estimation tasks. Based on ranking approach, Chen *et al.* [19] developed a novel CNN-based architecture “ranking-CNN” for age estimation. Zhang *et al.* [20], proposed a new CNN-based method, “Residual Networks of Residual Networks (RoR)” for age and gender estimation in-the-wild. Later, Rothe *et al.* [21] used “Deep EXpectation” (DEX), a deep

learning solution that is based on VGG-16 architecture, to solve real age estimation from a single face image without the use of facial landmarks. The authors in [22] developed a CNN architecture based on the “multi-class focal loss function” to improve the achievement of age estimation. In 2019, Li *et al.* [23] proposed a CNN-based technique, BridgeNet, for age estimation comprising two components; local regressors and gating networks. Liu *et al.* [24] then developed a method that is an extension of their work in [22]. The work is an end-to-end ordinal deep learning (ODL) framework, including two ordinal regression loss functions; square loss and cross-entropy loss. Zhang *et al.* [25] also proposed a novel method; recurrent age estimation (RAE). The CNN-based method makes use of the appearance features and the personalized aging patterns of input face images. In 2020, Nam *et al.* [26] solved the problem of age estimation of low-resolution facial images, with a deep CNN-based model that reconstructs low-resolution faces as high-resolution faces. Further, in [5], the authors developed a lightweight CNN network (ShuffleNetV2), based on the mixed attention mechanism (MA-SFV2).

B. APPARENT AGE

Apparent age describes “how old a person looks like?”. A significant amount of study has been done to extract facial features from faces and determine the people’s apparent age. For instance, in [27], Rothe *et al.*, developed a classification-based solution (Deep EXpectation (DEX)) for apparent age. The authors used the VGG-16 architecture that was initially pre-trained on ImageNet before further fine-tuned on the newly collected IMDB-WIKI dataset. Liu *et al.* in [28] later presented an hybrid model (AgeNet) that fuse a regression (real-value) and classification (Gaussian label distribution), to solve the apparent age estimation task. The authors in [29] studied a method that utilized the deep representations trained in a cascaded way. The approach also employed GoogleNet design, initially pre-trained with face images without age labels, then on data with chronological age labels to fine-tune the network parameters. In 2015, Ranjan *et al.* [30] developed an automatic age estimation method that is regression-based. The approach estimates the apparent age from unconstrained images using deep Convolutional Neural Networks (DCNN). However, some outliers in the input data can cause a large error term, which can lead to an unstable training process. Also in [31], Huo *et al.* introduced a deep CNN with distribution-based loss functions. The distributions utilized the ambiguity induced via manual labeling by learning a better model rather than using ages as the target. However, a distribution-based loss function might yield an inconsistency result during the training stage. In [32], the authors proposed a pre-trained VGG-16 CNN model that combined two separate models: The general and the children. The general model was initially trained on the huge IMDB-WIKI dataset for biological age estimation and then fine-tuned for apparent age estimation task. The children model used a pre-trained VGG-16 network and trained the

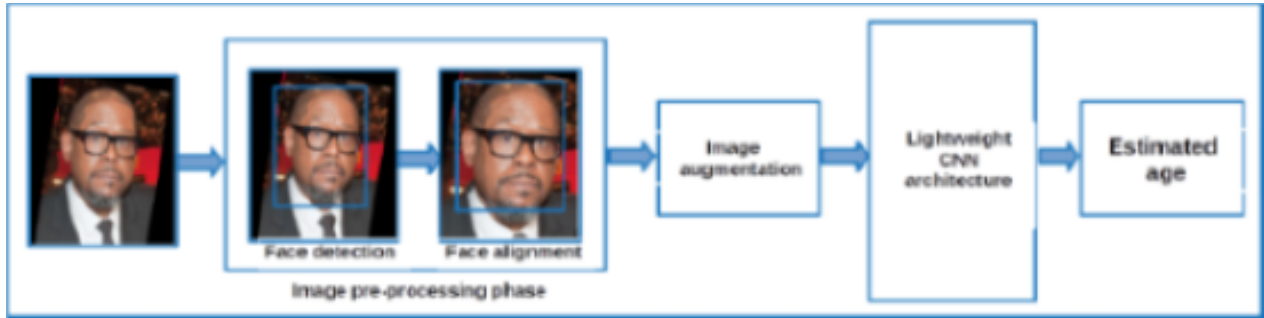


FIGURE 1. The pipeline of our proposed method.

children (private dataset of children between 0 and 12 years old) dataset on it. In [9], Agustsson *et al.* proposed a model: Residual DEX, an enhancement of DEX in [27]. The apparent age model was addressed as a classification problem that considers the age value as an independence category. The authors in [33] designed a lightweight CNN architecture that collectively learned age distribution and regressed it. The CNN-based approach, ThinAgeNet, employed the compression rate of 0.5. Rothe *et al.* [21] proposed Deep EXpectation (DEX) model that is based on VGG-16 architecture. The approach is a solution for apparent age estimation from a single face image without facial landmarks. In 2019, Li *et al.* [23] proposed a CNN based technique, BridgeNet, for apparent age estimation. The proposed model comprises two components; local regressors and gating networks that can jointly be learned in an end-to-end way. Also, in the same year, Liu *et al.* [24] developed a method that is an extension of their work in [22]. The work is an end-to-end ordinal deep learning (ODL) framework, including two ordinal regression loss functions; square loss and cross-entropy loss.

Although most of these methods, as mentioned above, improved classification performance of both real and apparent age, they employed heavyweight CNN networks such as ResNet, VGG-16, and GoogleNet, which incur expensive computation cost and storage overhead and inappropriate on resource-constrained mobile devices. Also, some of those methods can not achieve higher accuracy on unconstrained face images. Therefore, we propose a lightweight network with fewer parameters than those heavyweight networks that significantly improved the deployment process on mobile terminals. The proposed model also improved classification accuracy on unconstrained real-time images.

III. PROPOSED METHOD

Our proposed method for the age estimation task follows the pipeline described in Figure 1. A detailed description of each step is explained in the following subsections:

A. FACE DETECTION AND ALIGNMENT

Some of the faces in the dataset were taken under different imaging conditions, which makes it more challenging for

TABLE 1. The overall architecture of our lightweight network.

Layer type	Output size	Filter size/Stride
Input	$224 \times 224 \times 3$	—
Conv1	$56 \times 56 \times 96$	$7 \times 7/4 \times 4$
Stage2	$28 \times 28 \times 256$	5×5
Stage3	$14 \times 14 \times 384$	3×3
Conv4	$7 \times 7 \times 384$	3×3
HPooling		
Softmax		

recognition, as such, we prepare and preprocess the face images for training and testing stage.

1) FACE DETECTION

For this task, we employed a face detector algorithm described by Mathias *et al.* [34], to output bounding boxes of faces in the images. In a case, the image contains more than one face; the model will choose the faces with the highest detection scores. Further, we address faces that appear at the border and corner of the images by adding zero paddings and create square sized images. The face detector is very effective since it does not have about 2% of the images whose faces are not detected.

2) FACE ALIGNMENT

For face alignment, we implement the approach in [35], which initially detects the 68 facial landmark and then calculates the five points on the face. Finally, we align the faces to those points and then crop and resize to 224×224 .

B. CNN ARCHITECTURE

In this section, we describe our lightweight architecture of the convolutional neural network for real and apparent age classification. In designing our CNN, we consider the model size and efficiency. The network is a modification of the CNN architecture in [33], [36].

Firstly, we reduce the number of filters in each convolutional layer to reduce the model size and make it thinner. We also replace the fully-connected layer with a lighter but

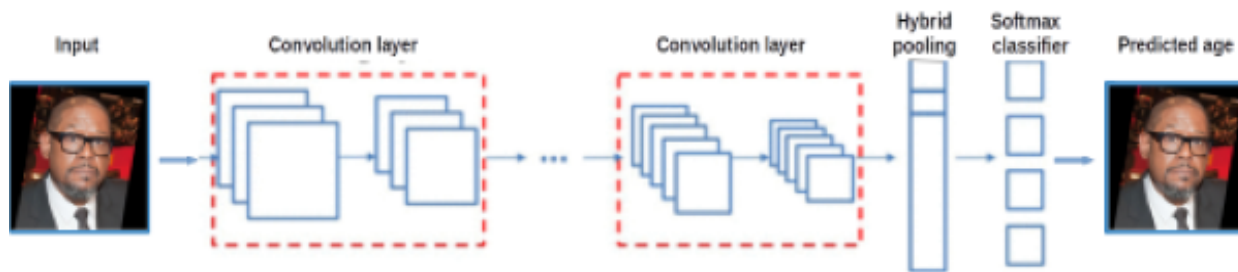


FIGURE 2. The composition of our lightweight convolutional neural networks architecture.

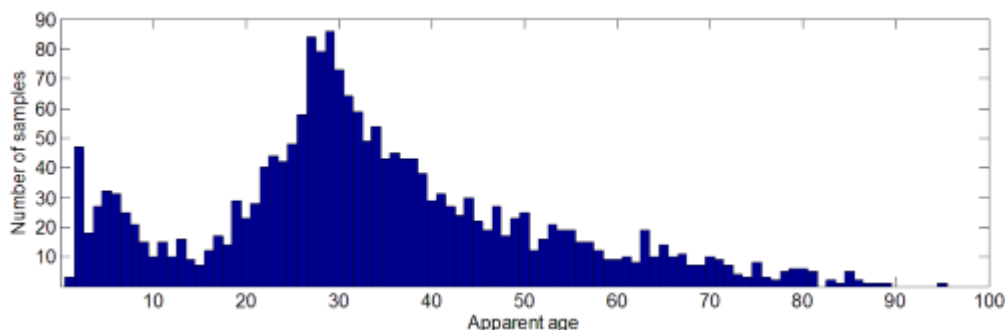


FIGURE 3. Number of image samples in APPA-REAL database per apparent age.

more efficient parameter (hybrid-pooling). Further, to speed up training, we replace the conventional Local Response Normalization (LRN) with batch normalization [37] and add it at the end of each CONV layer to improve the training speed. We model the real and apparent age estimation task as a deep classification problem, we therefore add a multi-class module after the Hybrid pooling, as shown in Figure 2. The novel CNN is a demonstration of a much faster and more accurate network.

Image augmentation. In this work, we employed FG-NET, MORPH-II and APPA-REAL datasets. FG-NET contains 1002 images, MORPH-II with 55,134 images while APPA-REAL has 7591 face images with a default split of 4113 for training, 1500 for testing, and 1978 for validation. However, the training set of these datasets is too low to train on any deep CNN architecture. Also, the age distribution of these database benchmark is uneven (see Figure 3). To address this, we need to employ an adaptive augmentation (regularization) model that increases the number of training images, and also make the age distribution of the training set even. The model also handles some imaging conditions, including illumination, diverse backgrounds, face position, image color, and quality. The regularization solution includes random scaling, random horizontal flipping, color channel shifting, standard color jittering, and random rotation. This generates an alter copies of every training image, for the network to take a different variation of the original image.

IV. EXPERIMENTS

A. DATASET

We conduct experiments to validate the effectiveness of the proposed method on two types of facial aging datasets. The first type contains two real age datasets with actual age labels which are collected in-the-wild. The second type is a dataset containing both real and apparent age labels. We present a brief description of the datasets below:

FG-NET (Face and Gesture Recognition Research Network Aging) database [38] is a publicly-available aging dataset collected in the year 2004 to support research work in human age estimation. The database has 1002 face images of 82 different individuals with ages ranging between 0 and 69 years. The database contains facial images that were produced through the scanning of photographs of different subjects. However, these collections reflect some significant variability in image resolution, expression, and illumination, among others. The images showed some differences in the size, background, tone, and lighting conditions of the images with some degree of occlusions of different forms in their appearance.

MORPH-II (Craniofacial Longitudinal Morphological Face) database [39] is a publicly-available facial aging database with a total of 55,134 facial pictures of more than 13,000 people collected at the University of North Carolina by the face aging group. Each of the pictures is marked with information such as the age, gender, ethnicity, weight, height, and ancestry of the character.



FIGURE 4. Some of the image samples in FG-NET, MORPH-II and APPA-REAL dataset.

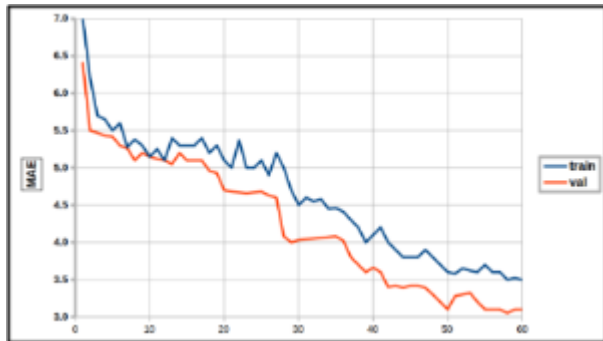


FIGURE 5. MAE (Mean Absolute Error) curves of the experiments on FG-NET dataset.

APPA-REAL database [9] is the first state-of-the-art database with both real and apparent age labels. The images are collected using labeling application, "crowd-sourcing" data collection, data from the "AgeGuess platform," and with the assistance of "Amazon Mechanical Turk" (AMT) workers. APPA-REAL database contains a total of 7591 images with the real and apparent age annotations. It has an age range that is between 0 and 95 of images of subjects that were taken under different conditions, which makes it more challenging for recognition purposes. Some examples of the images are shown in Figure 4. while the number analysis for APPA-REAL dataset is shown in Table 2.

TABLE 2. Number of the APPA-REAL images.

Part of Dataset	Number of Images
Training	4113
Testing	1978
Validation	1500
Total	7591

B. EVALUATION METRICS

At present, the mean absolute error (MAE) is one of the standard evaluation metrics in the research of real and apparent

age estimation [9]. It is defined as the average of the absolute deviations of all age estimates and actual values. Although another metric called normal-error (ϵ -error) employed in the apparent age estimation was recorded in [40], which considers the ground truth standard deviation, MAE can be used for both apparent and real age estimation. In this work, we, therefore, report the quantitative results of the experiments for real and apparent age in terms of Mean Absolute Error (MAE).

$$MAE = \sum_{i=1}^N \frac{|l_i - l_i^*|}{N} \tag{1}$$

where

- l_i : the estimated age
- l_i^* : the ground truth age for the test image i
- N : the total number of test images.

C. EXPERIMENT ON DIFFERENT DATASETS

1) EXPERIMENT ON FG-NET DATASET

We use the same image pre-processing pipeline as in all compared methods that include face detection, landmark detection, and face alignment. Since FG-NET has a total of 1002 images of 82 individual pictures, therefore we applied data augmentation to the training images of the dataset that produce alter copies of every training images and increase the images for effective training on our CNN architecture. The process produces a different variation of the original image before feeding the outcome into the novel lightweight convolution neural network architecture. These multiple sets of empirical experiments are conducted under the same data splits common in the literature: each specific analysis using all age images of one person as the test set, and all other remaining images as the training set. We finally achieved an MAE of 3.05 after a series of empirical experiment (with and without data augmentation) on the training images.

2) EXPERIMENT ON MORPH-II DATASET

There is no train/test split provided in the MORPH-II dataset, 10-fold cross-validation is the most common protocol in the literature; therefore, we divided the datasets into three parts: training, validation, and test set. 70% of the datasets were used for training, 20% for validation, and the remaining for testing. We also conducted experiments with different variations of the network and regularization algorithm to know the impact of each of them. It can be observed after the multiple set of the experiments that the lightweight CNN architecture with image augmentation, and the combined use of classification loss function, achieved the desirable final MAE result that is relatively low. This indicates that the novel method proposed in this work contributes to improving the accuracy of age estimation with an MAE of 2.31. The final MAE was achieved by experimenting with the test images of the dataset on the newly-trained model to validate the performance of the model.

3) EXPERIMENT ON APPA-REAL DATASET

We present experiments of the real and apparent age estimation conducted on the APPA-REAL dataset, which consists of 4113 for training, 1500 for testing, and 1978 images for validation.

Before training the network to predict the real and apparent age, we initially detected the face bounding boxes from the original images with the face detection model described in section III (A). We also detect the 68 facial landmark to calculates the five points on the face and finally align the faces to those points and then crop and resize to the appropriate shape. This is important so that the aligned faces are fed to the designed networks to extract the face descriptors needed for the estimation task.

Furthermore, to generate an alter copies of every training image, it is necessary to notify that all face images were augmented by random scaling, random horizontal flipping, color channel shifting, standard color jittering, and random rotation. Also, in our experiments, to learn face representation for ages, we pre-trained our multi-task lightweight CNN model from scratch using different optimization hyper-parameter. Some of the parameter settings employed in our training are optimizer (SGD), momentum (0.9), initial learning rate (0.001), learning rate decay(0.0005), etc. The whole training process was converged until the validation error remained minimal and constant. However, during the testing stage, we directly fed the testing facial images to the trained network to obtain the final age estimation.

All experiments are implemented on LENGAU CLUSTER: Intel® Xeon® 2.6 GHz; Memory 148.5 TB. The training of each experiment took about 5 hours. The software employed for the implementation includes:

- Ubuntu
- Python3
- keras
- scipy, numpy, Pandas, tqdm, tables, h5py
- OpenCV3

D. QUANTITATIVE COMPARISON WITH OTHER METHODS

1) COMPARISON ON FG-NET DATASET

Table 3 presents the mean average error (MAE) results of the latest facial age estimation methods including DEX, GAN, ODL, GA-DFL, AgeNet and MA-SFV2 on the FG-NET dataset. It can be deduced from the table that the MAE value of our method achieves an MAE of 3.05 (see figure 5) when only training the CNN model on the training set of FG-NET images. The result was significantly better than the other deep learning-based age estimation models presented in the table, demonstrating that our method outperforms the state-of-the-art methods with a clear margin on datasets. However, our result is only comparable to the performance obtained by the work of Wanhua *et al.*. The latter pre-trained its model on an external and large-scale dataset that contain more than half a million labeled images which added to the network model. In our case, only the original and single dataset is used.

2) COMPARISON ON MORPH-II DATASET

The performances of some state-of-the-art approaches are presented in Table 4. As can be seen, our lightweight CNN method achieves the lowest MAE of 2.31 (see figure 6)

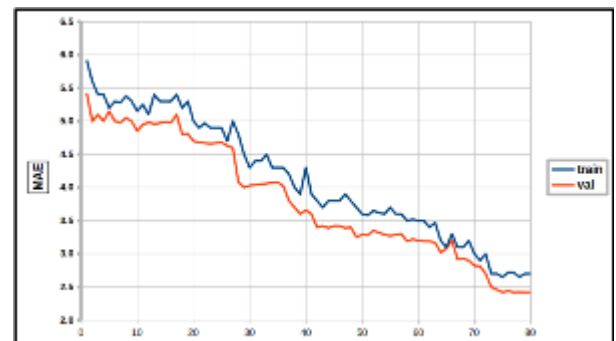


FIGURE 6. MAE (Mean Absolute Error) curves of the experiments on MORPH-II dataset.

TABLE 3. Performance comparison of different methods for age estimation on FG-NET dataset.

Methods	FG-NET
GA-DFL [41]	3.93
DEX [21]	3.09
AgeNet + divide-&-rule [42]	4.02
ODL(cross-entropy) [24]	3.71
BridgeNet [23]	2.56
GAN [26]	8.56
MA-SFV2 [5]	3.81
Proposed ¹	3.56
Proposed ²	3.05

without data augmentation: ¹ with data augmentation: ²

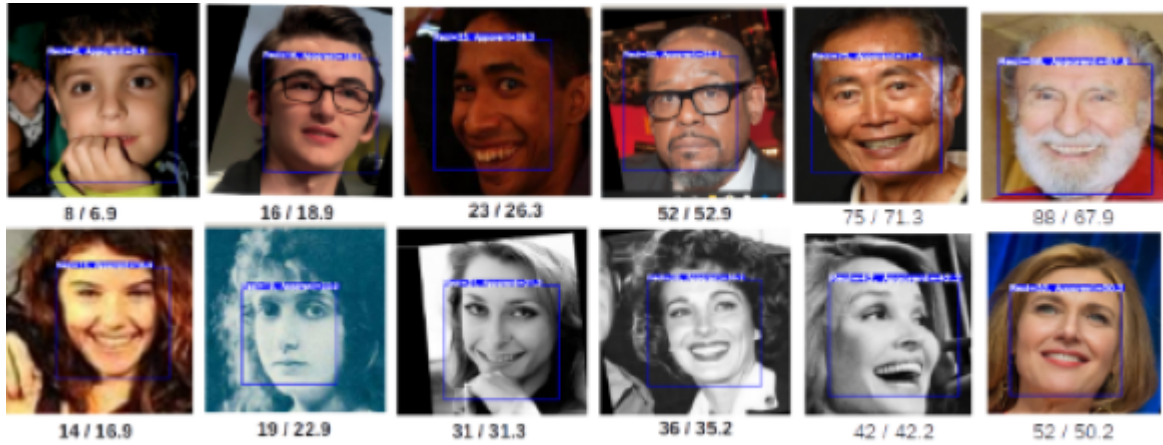


FIGURE 7. Visualization of the real and apparent age estimates on sample face images.

TABLE 4. Performance comparison of different methods for age estimation on MORPH-II dataset.

Methods	MORPH-II
AgeNet [28]	3.33
OR-CNN [43]	3.27
MRCNN [44]	3.48
GA-DFL [41]	3.25
ELM [45]	3.44
RAGN [46]	2.61
AgeNet + divide-&-rule [42]	3.48
DEX [21]	2.68
BridgeNet [23]	2.38
GAN [26]	9.42
MA-SFV2 [5]	2.68
Proposed ¹	2.72
Proposed ²	2.31

without data augmentation: ¹ with data augmentation: ²

on the MORPH-II dataset, which confirms the age estimation method proposed in this work had a good impact on the dataset. Although, the CNN-based methods like DEX, AgeNet, and RAGN, were pre-trained on a large amount of training data such as IMDB-WIKI, CACD, during the training process; they are not as effective as our method, with a better result than those deep learning models. Also, the network model of those methods is huge and time-consuming to train. It takes up a lot of storage space compared with our lightweight network model, which has a short training time, occupies a small amount of space, and ease of deployment on mobile devices.

3) COMPARISON ON APPA-REAL DATASET

Table 5 shows the MAE results of state-of-the-art age estimation methods such as ThinAgeNet and residual DEX on the APPA-REAL dataset. It can be seen from the comparison that the MAE result for age (real and apparent) estimation method proposed in this work, had a good effect achieving a comparable result when compared with the existing state-of-the-art age estimation methods.

TABLE 5. Performance comparison of different methods for age estimation on APPA-REAL dataset.

Methods	MAE (Real)	MAE (Apparent)
ThinAgeNet	-	3.452
Residual DEX	5.296	4.080
Proposed ¹	5.315	4.192
Proposed ²	4.942	3.809

without data augmentation: ¹ with data augmentation: ²

Further, our lightweight network model has a lighter model size with short training time. In contrast, the network model of some of those existing deep learning-based methods is huge, with extended training time, a huge training dataset, and ample storage space. This makes our model convenient for deployment on mobile devices with the resource-constrained facility. Some of the results of the classification model is presented in Figure 7.

4) COMPUTATIONAL TIME

The open-source deep learning software was employed in our model. We trained our CNN-based model using Intel® Xeon® 2.6 GHz with a memory of 148.5 TB. Table 6 presents the comparisons of the computational time during the testing phase and the model parameters of some previous CNN models. From the table, VGGNet took an

TABLE 6. Comparisons of computation time and model parameters of different CNN architectures on MORPH-II dataset.

Architecture	Testing time(imgs/s)	#Parameters(M)
VGGNet [21]	143.2	134.6
AlexNet [47]	2425.3	58.3
ResNet [48]	256.8	60.2
Proposed	41.05	2.7

average of 143.2 images per second on the employed cluster system. It also takes an average of 2425.3img/s and 256.8img/s computation time on AlexNet and ResNet, respectively. Moreover, the lightweight model parameter of 2.7M employed in our experiments takes 41.05 images per second during the testing phase. This proves that our model achieves the best performance with fewer parameters, and a reduction in computational time which satisfies the real-time requirement of the model for its applicability on mobile devices.

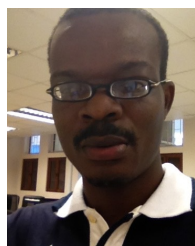
V. CONCLUSION AND FUTURE WORK

This paper was based on a lightweight convolutional neural network, combining robust image pre-processing algorithm and adaptive image augmentation. The model still achieved higher age estimation accuracy results that is comparable with the current state-of-the-art on FG-NET, MORPH-II and APPA-REAL datasets despite the lighter CNN model design with low training time and not-too large training images. For future works, researchers need to develop lighter CNN models with fewer parameters to continue to explore how to use a more compact model to achieve the desired age estimation result for its applicability and deployment on a mobile terminal. Also, there is need to consider a more robust and quality image pre-processing algorithm that detect the unfiltered images faster for real-time estimation of the images. Collection of non-frontal face images will also ease the problem age classification of unfiltered faces.

REFERENCES

- [1] K. A. Drobnyh and A. N. Polovinkin, "USING SUPERVISED DEEP LEARNING FOR HUMAN AGE ESTIMATION PROBLEM," *ISPRS-Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vols. 2-4, pp. 97-100, May 2017.
- [2] I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, "A deep analysis on age estimation," *Pattern Recognit. Lett.*, vol. 68, pp. 239-249, Dec. 2015.
- [3] A. R. Abbas and A. R. Kareem, "Intelligent age estimation from facial images using machine learning techniques," *Iraqi J. Sci.*, vol. 59, no. 2A, pp. 724-732, 2018.
- [4] J. Ruiz-del-Solar, R. Verschae, and M. Correa, "Recognition of faces in unconstrained environments: A comparative study," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1-5, Dec. 2009.
- [5] X. Liu, Y. Zou, H. Kuang, and X. Ma, "Face image age estimation based on data augmentation and lightweight convolutional neural network," *Symmetry*, vol. 12, no. 1, p. 146, Jan. 2020.
- [6] I. Bouchrika, N. Harrati, A. Ladjailia, and S. Khedairia, "Age estimation from facial images based on hierarchical feature selection," in *Proc. 16th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*, Dec. 2015, pp. 393-397.
- [7] O. F. W. Onifade and N. Department of Computer Science University of Ibadan, "A review on the suitability of machine learning approaches to facial age estimation," *Int. J. Mod. Edu. Comput. Sci.*, vol. 7, no. 12, pp. 17-28, Dec. 2015.
- [8] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. Yuille, "Deep regression forests for age estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2304-2313.
- [9] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, and R. Rothe, "Apparent and real age estimation in still images with deep residual regressors on appa-real database," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 87-94.
- [10] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3476-3483.
- [11] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, Jan. 2014, pp. 1988-1996.
- [12] M. Zulfiqar, F. Syed, M. J. Khan, and K. Khurshid, "Deep face recognition for biometric authentication," in *Proc. Int. Conf. Electr., Commun., Comput. Eng. (ICECCE)*, Jul. 2019, pp. 1-6.
- [13] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatio-spectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127-1137, Mar. 2015.
- [14] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701-1708.
- [15] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 34-42.
- [16] A. Yousaf, M. J. Khan, M. J. Khan, A. M. Siddiqui, and K. Khurshid, "A robust and efficient convolutional deep learning framework for age-invariant face recognition," *Expert Syst.*, vol. 37, no. 3, 2020, Art. no. e12503.
- [17] H. Liu, J. Lu, J. Feng, and J. Zhou, "Label-sensitive deep metric learning for facial age estimation," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 2, pp. 292-305, Feb. 2018.
- [18] A. Rosebrock, *Deep Learning for Computer Vision With Python*. 2018.
- [19] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-CNN for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 742-751.
- [20] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492-22503, 2017.
- [21] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, vol. 126, nos. 2-4, pp. 144-157, Apr. 2018.
- [22] W. Liu, L. Chen, and Y. Chen, "Age classification using convolutional neural networks with the multi-class focal loss," in *Proc. IOP Conf., Mater. Sci. Eng.*, 2018, vol. 428, no. 1, pp. 1-7.
- [23] W. Li, J. Lu, J. Feng, C. Xu, J. Zhou, and Q. Tian, "BridgeNet: A continuity-aware probabilistic network for age estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1145-1154.
- [24] H. Liu, J. Lu, J. Feng, and J. Zhou, "Ordinal deep learning for facial age estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 486-501, Feb. 2019.
- [25] H. Zhang, X. Geng, Y. Zhang, and F. Cheng, "Recurrent age estimation," *Pattern Recognit. Lett.*, vol. 125, pp. 271-277, Jul. 2019.
- [26] S. H. Nam, Y. H. Kim, N. Q. Truong, J. Choi, and K. R. Park, "Age estimation by super-resolution reconstruction based on adversarial networks," *IEEE Access*, vol. 8, pp. 17103-17120, 2020.
- [27] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep EXpectation of apparent age from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 252-257.
- [28] X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, and X. Chen, "AgeNet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 258-266.
- [29] Y. Zhu, Y. Li, G. Mu, and G. Guo, "A study on apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 267-273.

- [30] R. Ranjan, S. Zhou, J. C. Chen, A. Kumar, A. Alavi, V. M. Patel, and R. Chellappa, "Unconstrained age estimation with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 351–359.
- [31] Z. Huo, X. Yang, C. Xing, Y. Zhou, P. Hou, J. Lv, and X. Geng, "Deep age distribution learning for apparent age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 722–729.
- [32] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 801–809.
- [33] B.-B. Gao, H.-Y. Zhou, J. Wu, and X. Geng, "Age estimation using expectation of label distribution learning," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 712–718.
- [34] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool, "Face detection without bells and whistles," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 8692. Cham, Switzerland: Springer, 2014, pp. 720–735.
- [35] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.
- [36] O. Agbo-Ajala and S. Viriri, "Deeply learned classifiers for age and gender predictions of unfiltered faces," *Sci. World J.*, vol. 2020, pp. 1–12, Apr. 2020.
- [37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, pp. 448–456, 2015.
- [38] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the FG-NET ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, Jun. 2016.
- [39] K. Ricanek, Jr., and T. Tesafaye, "MORPH: A longitudinal image Age-progression, of normal adult," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 1–4.
- [40] S. Escalera, J. Fabian, P. Pardo, X. Baro, J. Gonzalez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon, "ChaLearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 243–251.
- [41] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.
- [42] H. Liao, Y. Yan, W. Dai, and P. Fan, "Age estimation of face images based on CNN and Divide-and-Rule strategy," *Math. Problems Eng.*, vol. 2018, pp. 1–8, Jun. 2018.
- [43] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output CNN for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4920–4928.
- [44] K. H. Liu, S. Yan, and C. C. Kuo, "Age estimation via grouping and decision fusion," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2408–2423, Nov. 2015.
- [45] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN-ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, Jan. 2018.
- [46] M. Duan, K. Li, and K. Li, "An ensemble CNN2ELM for age estimation," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 758–772, Mar. 2018.
- [47] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.



OLATUNBOSUN AGBO-AJALA received the degrees from the University of Ibadan, Nigeria, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Sciences, University of Kwazulu-Natal, South Africa. He has published articles in both local and international journals, and conferences. His research interests include computer vision, image processing, and deep learning.



SERESTINA VIRIRI (Member, IEEE) is currently working as a Professor of computer sciences with the School of Mathematics, Statistics, and Computer Science, University of Kwazulu-Natal, South Africa. He has graduated several M.Sc. and Ph.D. students. His research interests include computer vision, image processing, pattern recognition, and other image processing related areas, such as biometrics, medical imaging, and nuclear medicine. He is an NRF-Rated Researcher. He has published extensively in several accredited *Computer Vision* and *Image Processing* journals, international, and national conference proceedings. He serves as a Reviewer for several accredited journals. He has served on program committees for numerous international and national conferences.

...