Facial Age Estimation Using Convolution Neural Networks

Abdulfattah E. Ba Alawi
Department of Software Engineering
Taiz University
Taiz, Yemen
baalawi.abdulfattah@gmail.com

Ahmed Y. A. Saeed
Department of Software Engineering
Taiz University
Taiz, Yemen
ahmed12.12f@gmail.com

2021 International Congress of Advanced Technology and Engineering (ICOTEN) | 978-1-6654-1224-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICOTEN52080.2021.9493508

***Abstract*— Age-related analysis has been a concern in the current years because many implementations have a great significance. The techniques of facial age prediction and classification are commonly used in the recent years for vitality applications but these techniques are time-consuming. The deep algorithms demonstrated superior efficiency compared to other approaches in order to solve the problem of age estimation. Herein, an age classification model is proposed using the mechanisms of deep age estimation in this article. This work introduces an age recognition model that helps to classify a person's image into a suitable aging group. The proposed model achieved better results in the prediction process with an accuracy reached 85.7% using InceptionV4. The main difference between this work and the relevant related works is that this work focuses on highlighting the performance of four pre-trained models three of them have different architectures; ResNet50, ResNet101, Sequeeze1_0, and InceptionV4. In deep age evaluation schemes, we look at previous study initiatives and current common datasets.**

Keywords—Age estimation, Prediction, age grouping, facial aging, CNN.

I. INTRODUCTION

Age-related analysis has been a concern in the current years because many implementations have a great significance. Age projections are used in access control schemes where a user's age support is offered. Tobacco vendors in Japan, for example, are prohibited from purchasing cigarettes by children and adult people under 20 years of age [1]. Marketing challenges include, for instance, the use of age-based ABCI schemes where publicity shows are based on the customer's age. All businesses must determine the age of their clients. Clients are categorized into categories by age. This will support businesses with the required designation promotion of their goods. In the protection sector, ABCI should be used even though the child is wrongly left alone if the car makes a warning. ABCI can also prohibit children under the age of protection from playing a risky game in a theme park [2]. Age estimates will also be used in bars to spot minors attempting to drink. In population studies for crime prevention, the age prediction system is still used. The age of the perpetrator identified by the security camera can be determined. Robotic nurses may take advantage of age analysis in health systems to accelerate the initial assistance [1].

Human aging science has lately become very involved in its widespread role in applications for vitality. Three significant areas are focused on aging studies. Next, the Age Invariant Facial Recognition[3] is an aid to recognize faces independently of their age changes (faces). In the real world, it has been used for passport or driving license regeneration and biometric authentication [3]. Secondly, Age synthesis, the method of forecasting the possible presence of a visual picture at a certain age[2]. Age synthesis can be used in forensic work to identify suspects.

A new method to extract functions based on deep learning has been investigated by Wang et al.[4]. The functionality and assistance of vector regression (SVR) were extracted using CNN to learn the pattern of aging. They could not refer to the impact of sex and race, where only knowledge about age was included. In [5] the large data set, MORPH Album 2, was reduced by Chang and Chen to 3,74 years and outperformed their methods. The CSOHR method, which is based on ordinary regression, was suggested. The variations in neutral facial expressions that have an effect on predicted age were not considered.

Indeed, as some recent reports have demonstrated, the aging development between men and women is different[6]. For one, wrinkles occur more and earlier in women than men. Before estimating age, recent works appear to decide gender. In [7], the authors used deep faith networks to prevent overfitting issues during the training of help classifiers train ideally. Next, they matched the facial picture to establish the sex and age of each sample before collecting features. They used the unfiltered pictures in an uncompromising atmosphere in the viewer dataset. A value of 95.3% of accuracy on Adience dataset was reached.

In 2016, during training a deep network, Dong et al.[8] used the transfer learning technique in predictive unlabeled images. The age labels are categorized and organized by using deep networks, and their relationships are dealt with through a loss function. The loss feature uses only the age gap information to estimate the age of the profound network, irrespective of race and gender. In [9], the authors overcome the absence of correct labeling with the turn of categorical labels into distinct age and sex labels. It suggested a multi-task approach to profound learning. The chance of picture I is of the age a1 they measured given that the gender form is g1 [10-12].

Nam et al. [13] used a CNN model to rebuild low-resolution faces into high-resolution faces to Deal with the problem of age estimate in low-resolution facial pictures. A conditional GAN pre-processed low-resolution face pictures before being utilized as input in the CNN method.

Depending on the mixed attention mechanism, Liu et al. [14] built a lightweight Deep Neural Network Model (ShuffeNetV2). This model transforms the output layer into a classification problem (which classifies age as a distinct category), a regression problem (which ranks the age of the facial expression in a certain order), and distribution learning.

Latest reports lead to modifications in the aging appearance of individuals as a result of discrepancies in the distortion of individuals and emotional expressions [6]. In image processing, the consistency, illumination, pose, and texture of face images can vary. These difficulties can be summed up as follows:

- The calculation of age depends on several different variables including gender, race, wrinkles... etc. These influences may be inherent or foreign. Sex, race, lifestyle, etc. can be intrinsic variables. Extrinsic influences can include ultraviolet sun rays, infections, changes in temperature, etc. Aging is a distinct and uncontrollable phase from person - to - person.
- Makeup, beauty treatments, and cosmetic operations may remove or minimize signs of aging on the face. This makes it more difficult to determine the age.
- A broad representative dataset with specific labels of images is challenging to search. The posture, light, consistency, and texture of facial pictures should be small and the race, sex, and age ranges should be narrowly modified in order to become more expressive. There should also be a fair amount of face images per user to make the dataset more profitable.
- It is difficult to explain aging periods in the same person, except through the internet, with a sufficient number of photos.

This work introduces an age recognition system that helps classify a person's image into suitable aging groups. The proposed model achieved better results in the prediction process with an accuracy reached 85.7%. The main difference between this work and the relevant related works is that this work focuses on highlighting the performance of four pre-trained models three of them have different architectures; ResNet50, ResNet101, Sequeeze1_0, and InceptionV4.

The remaining parts of the manuscript are organized in the following way: the following section illustrates the method used with a brief description of the techniques that have been implemented. Sect 2 discusses the performance of the four pre-trained models. The conclusion is presented in Sect 3.

II. Materials and Method

In this paper, age recognition based on facial images has been built. First, we collected the dataset from the internet and we did the preprocessing step for this dataset. and face detection step is used for detecting faces from images and the last step was for building the model by using deep learning algorithms.

A. Data Collection

We downloaded The dataset (UTK dataset) of this system from the website [19]. This dataset contains images for different people with ages from 0 to 111. The number of images of this dataset is 9780 which is divided into three-part, 70% for training, 20% for validation, and 10% for testing. for building this model, we divided the dataset into three classes, the first class has aged with a range from 0 to 17, the second class from 18 to 36, and the last class contains ages bigger than 36. The following table describes the content of the dataset.

TABLE I. Dataset Details

Class	Age	The number of samples
Class 0	From 0 to 17	3925
Class 1	From 18 to 36	2469
Class 2	> 36	3386

B. . Face Detection

In this phase, we used the Haar Cascade classifier to detect faces from images. the phase of face detection is shown in the following figure.

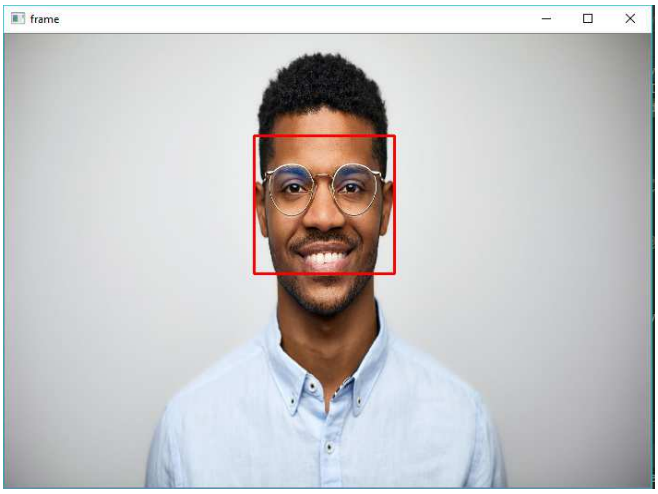

Fig. 1. Face detection step.

C. Data pre-processing

In this phase, we resized images into 224x224x3 in order to make images' dimensions the same as the input layer of pre-trained models. However, InceptionV4 has different dimensions (not 224x224x3 but 299x299x3), the images resize also in that dimensions.

D. Retraining CNN pre-trained models

In this phase, the dataset of this paper has been trained using pre-trained models which used four pre-trained models (ResNet-50, ResNet-101, Inception-V4, and SqueezeNet-1.0).

E. Pre-trained Models

a) ResNet Model

ResNet50 and ResNet101 are versions of Residual Networks that have been proposed by He et al. in 2014[18]. ResNet18 demonstrates high precision and processing time efficiency. Inputs are retrieved by the encoder by multiple convergence variations, Corrected Linear Units (ReLU) [15]. The encoder In addition to the network does not include pooling layers. In comparison to the first overlay, the filter scale is 7 to 7 in the Resnet-50 network, and the other layers are three to three. We conduct the batch standardization (Batch Norm)[16] for each cooling layer on the encoder in order to accelerate the network training.

b) SqueezeNet

In the beginning, SqueezeNet was defined in the article "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size."[17] AlexNet is a deep neuronal network with 240 MB of parameters; the parameters in SqueezeNet are 5 MB. Nevertheless, it is important to remember that SqueezeNet is not a "squeezed version of AlexNet." SqueezeNet is an architecture very independent from AlexNet.

c) Inception V4

Inception-v4 is an involute architecture that expands upon previous versions of the Inception family by architecture simplification and the use of more components of inception than Inception-v3. The architecture of Inception V4 is shown in the following figure.

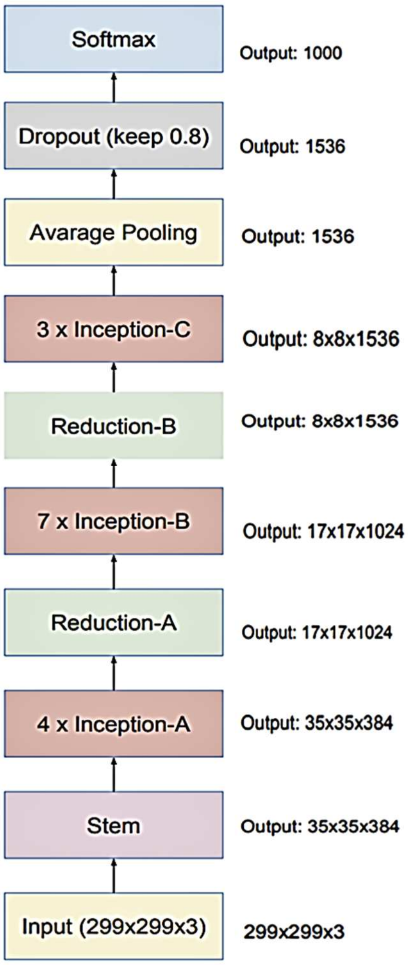

Fig. 2. InceptionV4 architecture.

III. RESULTS AND DISCUSSION

In this system, we trained the UKA dataset with 3 classes of Ages by using four pre-trained Models. the best training results were for Inception-v4 which the training loss achieved 0.281 and the accuracy achieved 91.60%, and the worst result was for ResNet-50 which the loss was 0.471 with an accuracy of 84.81%. the training Loss and accuracy of the pre-trained models are shown in the following figure (Fig. 3).

Fig. 3. Training Loss of the Four Pre-trained Models.

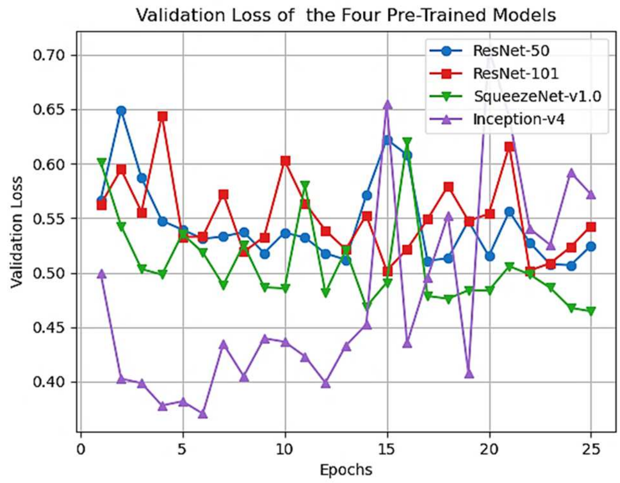

Fig. 4. Validation loss of the four pre-trained models.

As shown in the previous Figure (Fig. 4), The best validation results of the pre-trained models were for Inception which achieved 0.371 in Loss and the accuracy of the Model achieved 84.76%, and the worst results were for ResNet-50 which the Loss of the Model achieved 0.507 and the accuracy was 78.17%.

After the training phase, the test phase was implementing which used Accuracy, Recall, Precision, and F1_score to evaluate the performance of the pre-trained models. Table II. Show the performance of the pre-trained models.

TABLE II. THE PERFORMANCE OF THE PRE-TRAINED MODELS IN THE TESTING PHASE.

Model/Metrics	Loss	Accuracy	Recall	Precision	F1_score
ResNet-50	0.5024	0.7822	0.7467	0.7583	0.7489
ResNet-101	0.4921	0.8026	0.7815	0.7828	0.7820
SqueezeNet-V1.0	0.4801	0.8026	0.7907	0.7873	0.7850
Inception-V4	**0.3839**	**0.8579**	**0.8357**	**0.8384**	**0.8369**

As shown in the previous table. The best models in the testing phase were Inception-v4 which achieved in terms of Loss, Accuracy, Recall, Precision, and F1_score 0.3839, 0.8579, 0.8357, 0.8384, and 0.8369 respectively.

After this work had been done, we compared it with other previous works and the result was showing in the following Table.

The author	Year	Method	Result
Wang et al.[4]	2015	Deep Features , SVR, SVM, PLS and CCA	MAE= 4.26%
Yoo et all.[9]	2018	conditional multitask deep learning	MAE=2.91%
Liu et al. [14]	2020	Mixed Attention-ShuffleNetV2	MAE=2.68%
Our Proposed Models	2021	ResNet-50, ResNet-101, SqueezeNet-v1.0, Inception-V4	loss=3.8%, Acc=85.79%

To validate the system, we used some images to predict the ages of persons. The following figure shows the result of the prediction step.

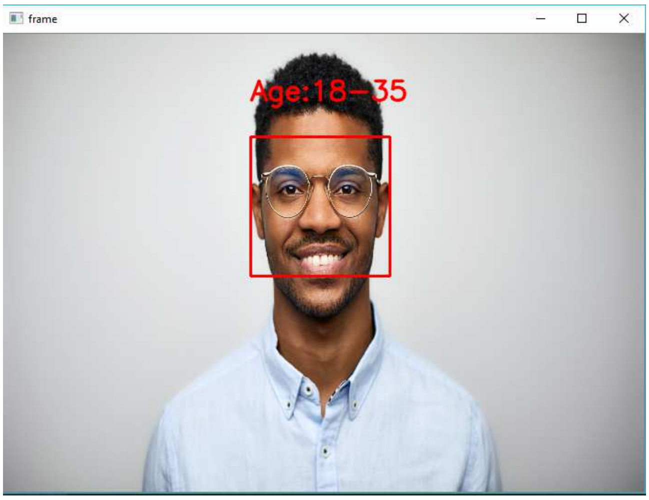

Fig. 5. Prediction of the age recognition system.

The above figure (Fig. 5) shows the corresponding class for the tested image.

IV. Conclusion

This paper presents a system for age recognition based on pre-trained models. We used four pre-trained models (e.g. ResNet-50, ResNet-101, squeezeNet-v1.0, and Inception-V4). The best model was Inception-v4 which achieved 0.3839 in Loss, 0.8579 in Accuracy, 0.8357 in Recall, 0.8384 in precision, and 0.8369 in F1_score. In the future, we plan to improve this model by grouping the dataset into more than four classes (e.g. child, young, adult, and old). Moreover, we plan to enhance the accuracy of recognition by using other models, in addition, to use different parameters to make models more stable.

Reference

[1] R. Angulu, J. R. Tapamo, A. O. J. E. J. o. I. Adewumi, and V. Processing, "Age estimation via face images: a survey," vol. 2018, no. 1, pp. 1-35, 2018.

[2] M. El Dib, "Automatic facial age estimation," Tesis de máster no publicada. Universidad del Cairo, Cairo, Egipto, 2011.

[3] M. M. Sawant and K. M. J. A. I. R. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," vol. 52, no. 2, pp. 981-1008, 2019.

[4] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in 2015 IEEE Winter Conference on Applications of Computer Vision, 2015, pp. 534-541: IEEE.

[5] K.-Y. Chang and C.-S. J. I. T. o. I. P. Chen, "A learning framework for age rank estimation based on face images with scattering transform," vol. 24, no. 3, pp. 785-798, 2015.

[6] K. Sveikata, I. Balciuniene, and J. J. S. Tutkuviene, "Factors influencing face aging. Literature review," vol. 13, no. 4, pp. 113-116, 2011.

[7] Y. Dong, Y. Liu, and S. J. N. Lian, "Automatic age estimation based on deep learning algorithm," vol. 187, pp. 4-10, 2016.

[8] E. Eidinger, R. Enbar, T. J. I. T. o. I. F. Hassner, and Security, "Age and gender estimation of unfiltered faces," vol. 9, no. 12, pp. 2170-2179, 2014.

[9] B. Yoo, Y. Kwak, Y. Kim, C. Choi, and J. J. I. S. P. L. Kim, "Deep facial age estimation using conditional multitask learning with weak label expansion," vol. 25, no. 6, pp. 808-812, 2018.

[10] C. N. Duong, K. G. Quach, K. Luu, H. B. Le, and K. Ricanek, "Fine tuning age estimation with global and local facial features," in IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, 2011, pp. 2032-2035.

[11] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in 2011 International Joint Conference on Biometrics (IJCB), 2011, pp. 1-6: IEEE.

[12] P. Koruga, M. Bača, and J. Ševa, "Application of modified anthropometric model in facial age estimation," in Proceedings ELMAR-2011, 2011, pp. 17-20: IEEE.

[13] S. H. Nam, Y. H. Kim, N. Q. Truong, J. Choi, and K. R. J. I. A. Park, "Age estimation by super-resolution reconstruction based on adversarial networks," vol. 8, pp. 17103-17120, 2020.

[14] X. Liu, Y. Zou, H. Kuang, and X. J. S. Ma, "Face image age estimation based on data augmentation and lightweight convolutional neural network," vol. 12, no. 1, p. 146, 2020.

[15] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines", *Proc. 27th Int. Conf. Mach. Learn.*, pp. 807-814, 2010.

[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv:1502.03167, 2015, [online] Available: https://arxiv.org/abs/1502.03167.

[17] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. J. a. p. a. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," 2016.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[19] UTKFace Dataset. (2021), available: susanqq.github.io/UTKFace (Accessed on January 30, 2021).