

# CNN Based Features Extraction for Age Estimation and Gender Classification

Mohammed Kamel Benkaddour

University Kasdi Marbah, Department of Computer Science and Information Technology

FNTIC Faculty, Ouargla, Algeria

E-mail: benkaddour.kamel@univ-ouargla.dz

**Keywords:** biometric, gender prediction, age estimation, deep neural network, convolutional neural networks (CNN)

**Received:** August 3, 2020

*In recent years, age estimation and gender classification was one of the issues most frequently discussed in the field of pattern recognition and computer vision. This paper proposes automated predictions of age and gender based features extraction from human facials images. Contrary to the other conventional approaches on the unfiltered face image, in this study, we show that a substantial improvement be obtained for these tasks by learning representations with the use of deep convolutional neural networks (CNN). The feedforward neural network method used in this research enhances robustness for highly variable unconstrained recognition tasks to identify the gender and age group estimation. This research was analyzed and validated for the gender prediction and age estimation on both the Essex face dataset and the Adience benchmark. The results obtained show that the proposed approach offers a major performance gain, our model achieve very interesting efficiency and the state-of-the-art performance in both age and gender scoring.*

*Povzetek: V prispevku je opisana študija s konvolucijskimi nevronskimi mrežami (CNN) za prepoznavanje starosti in spola iz obrazov na slikah.*

## 1 Introduction

Human face image analysis is an important research area in the field of pattern recognition and computer vision, in which many researchers concentrate on creating new or improving existing algorithms to several face perception tasks, including face recognition, age classification, gender recognition, etc. More closely, a face image is the source of many kinds of important information's, such as identity, expression, emotion, gender, race, age, etc. Human facial image processing research is undergoing in many directions and it still active and interesting, where age estimation and gender distinction from face images play important roles in many computer vision based applications as parental controls of the websites, video services and shopping recommendation systems [1].

Over the last decades, many methods have been proposed to tackle the age and gender classification task, the convolutional neural networks (CNN) is one of them more recently it has been employed in face image based age and gender classification tasks [2][3]. However, as face images vary in a wide range under the unconstrained conditions (namely, in the wild), the performances of CNN still need to be improved, especially in age estimation tasks [4]. In this study, we concentrate on the issue of age group classification rather than that of exact age estimation. Though extensive approaches for age classification have been proposed, most of them focused on constrained images such as FG-NET [7] and MORPH [8]. After training our model, we provide a significant

improvement in performance and enhance the recognition accuracy of age and gender classification.

The remainder of this paper is organized as follows: first briefly reviews the related work, then illustrates the proposed method and network architecture in training and testing of the data, after that the experimental results are given and finally, conclusions are drawn in.

## 2 Related works

The age and gender classification system is used to categorize human images into various groups determined by facial features. Accuracy of correctly classified image depending on the age, gender and their combination can be calculated by number of correctly classified images comparing to the stored values of the corresponding images in the database. Estimating human age group and gender automatically via facial image analysis has lots of potential real-world applications, such as human computer interaction and multimedia communication. Many researchers give their contribution to the human age and gender classification system which is a growing area in the research of the past decade [1][3].

The successful applications of CNN on many computer vision tasks have revealed that CNN is a powerful tool in image learning. If enough training data are given, CNN is able to learn a compact and discriminative image feature representation. Therefore, several studies have focused on the structuring of a new architecture of the CNN networks to obtain the best

recognition and prediction of age and gender classification [5] [6] [7]. In this part, we describe the previous works on age and gender classification.

Ari Ekmekji [9] Proposes a model for age and gender classification, where he modified first some effective architectures used for gender and age classification by reducing the number of parameters, increasing the depth of the network, and modifying the level of the used dropout rate. In addition, the second face of his project focuses on coupling the architectures for age and gender recognition to take advantage of the gender-specific age characteristics and age specific gender characteristics inherent to images. Antipov et al [5], in this work they design the state-of-the-art of gender recognition and age estimation models according to three popular datasets, LFW, MORPH-II and FG-NET, where they analyzed four important factors of the CNN for gender recognition and age estimation: (1) the target age encoding and loss function, (2) the CNN depth, (3) the need for pretraining, (4) the training strategy: mono-task or multi-task.

Gil Levi and Tal Hassaner [3] suggested a structure of CNN constructed from five layers, two fully-connected layers and three convolutional layers. The FERET dataset of images was used in the training and test samples. In this structure, two methods were applied in the prediction process : the first was center crop for cropping the source image to 227\*227 in center image , and the second was over-sampling, segmenting the image into five regions each one making 256\*256. Experimentally, they presented good results in gender and age prediction compared to the previous scholars.

### 3 Proposed method

#### 3.1 Convolutional neural networks

In neural networks, convolutional neural network is one of the main categories used for images recognition, images classifications. In 1995, Yann LeCun and Yoshua Bengio introduced the concept of convolutional neural networks to classify the image successfully and for recognizing handwritten control numbers by Lenet-5 [2].

LeCun’s convolutional neural networks [11] are organized in a sequence of alternating two types of layers S-layers and C-layers, called convolution and sub-sampling with one or more fully connected layers (FC) in the ends.

Convolutional Neural Networks are a special kind of multi-layer neural networks, they follow the path of its predecessor neocognitron in its shape, structure, and learning philosophy [3]. Traditionally, neural networks convert input data into a one-dimensional vector [4], and they have the ability to perform both feature extraction and classification. The input layer receives normalized images with the same sizes [2].

Each input image will pass through a series of convolution layers with filters (Kernels), Pooling, fully connected layers (FC) and apply activation function to classify an object with probabilistic values. Figure 1

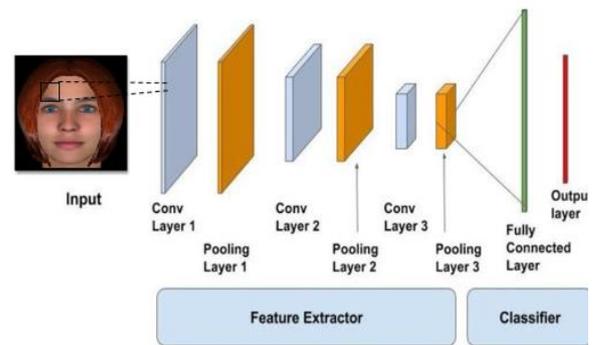


Figure 1: Convolutional network architecture.

shows a complete flow of CNN to process an input image and classifies the objects based on values [10].

The feature map of each pixel is calculated as follows:

$$C_n = f(x * W + b) \quad (1)$$

Where “\*” is the convolution computation, n the pixel in the feature map, x the pixel-value, W weights of the convolution kernel. f is a non-linear activation function like sigmoid or ReLu (Rectified Linear unit), applied to the result of the convolution.

After that some nonlinear function like the sigmoid or ReLu is applied to the result of this convolution referred as a feature map. The different feature maps obtained for different kernels are also called channels. Additionally, one also applies a pooling layer (or sub-sampling layer), where the output of the convolution with some filter is divided in tiled squared regions and each regions is summarized by a single value. This value can be the maximal response within the region in case of max-pooling or the average value of the region in case of mean-pooling, but most of the data scientists uses ReLU since performance is better than other activation function. The ReLU layer applies the function as follows:

$$f(x) = \max(0, x) \quad (2)$$

We used the ReLU function in our work. After ReLU function, we apply a pooling layer to reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or downsampling which reduces the dimensionality of each map but retains the important information. Spatial pooling can be of different types : max-pooling, average-pooling or sum-pooling. Max pooling take the largest element from the rectified feature map. Taking the largest element could also take the average pooling (sum of all elements in the feature map).

Convolutional layer and pooling layer compound the feature extraction part. Afterwards, it comes the fully connected layers which perform classification on the extracted features by the convolutional layers and the pooling layers. These layers are similar to the layers in Multilayer Perceptron (MLP). Finally we flattened our matrix into vector and feed it into a fully connected layer like neural network , we have an activation function such as softmax or sigmoid to classify the outputs. In the FC layer the neurons have complete connections with all the activations of the previous layer, as it is in in regular neural networks [11].

In the output layer, the activation function usually depends on the task, for example in the case of binary or multiple binary classification the sigmoid activation is used (which is guaranteed to have values in [0, 1]). For multi-way classification the softmax activation function is used to generate a value between 0–1 for each node.

### 3.2 Network architecture

The Architecture of our CNN as table1 shows ,is trained on a database to age and gender classification .In this section, we present our CNN model and explain the used methods in the project .

Name	Type	Description of output size
Input layer	Input image	3x92x112
Conv1	Convolution + ReLU	16 filters 3x3x3
S1	Max Pooling	16x2x2 stride 1
Conv2	Convolution + ReLU	32 filters 16x3x3
S3	Max Pooling	32x2x2 stride 1
Conv3	Convolution + ReLU	64 filters 32x3x3
S3	Max Pooling	64x2x2 stride 1
Con4	Convolution + ReLU	128 filters 64x3x3
S4	Max Pooling	128x2x2 stride 1
FC1	Fully Connected + dropout	5376x1x1
FC2	Fully Connected + dropout	512
FC3	Fully Connected + dropou	128

Table 1: Detailed architecture of our CNN used for features extraction.

The Shape of the image is 92 x 112 x 3 where 92 represents the height, 112 the width, and 3 represents the number of color channels. When we say 92 x 112 it means we have 10,304 pixels in the data and every pixel has an R-G-B value hence 3 color channels. CNN architectures require that all input images have the same resolution (height and width). In our work we used the padding on the convolutional layer with a default value to improves performance by keeping information at the borders. This means that the filter is applied only to valid paths to the entry.

The input images of our CNN pass through four convolution layers that form the network, each layer is followed by ReLU function with conv1 consist 16 filters , conv2 consist 32 filters , conv3 consist 64 filters and conv4 consist 128 filters , then four pooling layers are used layer S1, S2 , S3 and S4 are subsampling layers , after the preprocessing layers comes three fully connected layers each one of them is followed by a and dropout layer.FC1 with 5376 neurons , FC2 with 512 neurons and FC3 with 128 neurons, yield to normalized class scores for either gender or age, respectively.

Finally there is softmax layer that sits on the top of FC3, which gives the loss probabilities for the age estimation, the sigmoid layer it also sits on the top of FC3 and gives the loss class probabilities for the gender classification.

### 3.3 Training and testing

Over the last decade, several research works have been published on facial gender and age estimation. The algorithms usually take one of two approaches: age group or age-specific estimation. The former classifies a person as either child or adult, while the latter is more precise as it attempts to estimate the age group of a person. Each of these approaches can be further decomposed into two key

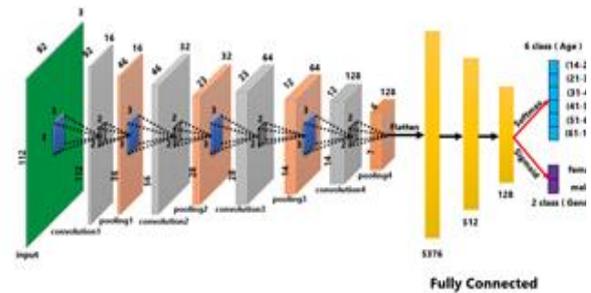


Figure 2: Our Convolutional neural network architecture.

steps: feature extraction and pattern learning classification.

In our project, we divided the dataset into 2 folders train data and test data , each person in the dataset has 10 images , we put 8 of them in the training part and 2 in the testing part . In the age estimation there are 6 classes (0-6), (8-20), (25-32), (38-43), (48-53) and (+ 60), the gender "m : male" "f: female" for the label we took from the name on the images so that we renamed all the pictures for example : "Adele-1-[f,(48-53)]", "dagran-2-[m,(38-43)]". So each photo in training and test data has two labels, gender label and age group label.

## 4 Datasets

In this section, we introduce the datasets used for the evaluation of our proposed approach with a description of their specifications for age group estimations and gender classifications accuracy.

### 4.2 Essex database

The Essex face data is a dataset of the university Essex , the images are taken with a plain green background , where the subjects sit at fixed distance from the camera and are asked to speak, whilst a sequence of images is taken. The speech is used to introduce facial expression variation.This dataset contains images of male and female subjects. The individuals are a students, so the majority of individuals are between (20- 32) years old but some older individuals are also present [15].

We have made some changes to the images of this data by changing the image format from JPEG to BMP and its size from (180\*200) to (92\*112). In our work , we take from 20 photos 10 photos for every subject (2 for the test data, 8 for the train data) and manually labeled these images. As an example, images corresponding to one individual are shown in the Figure 3.



Figure 5: Images from the Adience benchmark.



Figure 6: Sample images of one person from Essex database.

### 4.3 Adience benchmark

We train and test our deep convolutional neural network model on the adience benchmark. Adience dataset is a collection of face images from ideal real-life and unconstrained environments contain approximately 26 K images of 2,284 with different age group with the corresponding gender label. The Adience benchmark measures both the accuracy in gender and age and display a high-level of variations in noise, pose, and appearance, among others [16].

## 5 Experiments, results and discussions

In this section, we first introduce the parameter settings of the experimental analysis, then results of the experiments evaluated on both Essex face dataset and Adience benchmark for gender prediction and age estimation are given, finally we draw on a performance comparison of our results with state-of-the-art works.

### 5.1 Evaluation of results

The evaluation of the proposed methodology can be summarized using the classification accuracy rate and loss rate of each set of data.

$$\text{Accuracy rate} = \frac{\text{Correctly classified images in Subset}}{\text{Total Number of images in Subset}} \quad (3)$$

$$\text{Acc (total)} = \text{Accuracy(Age)} + \text{Accuracy(Gender)} \quad (4)$$

$$\text{Loss (total)} = \text{Loss(Age)} + \text{Loss(Gender)} \quad (5)$$

Where Accuracy of correctly classified images depending on the age or gender is calculated by comparing the results of the algorithm of classification and the stored age and gender values of the images in the images labels. After training and testing our model according to

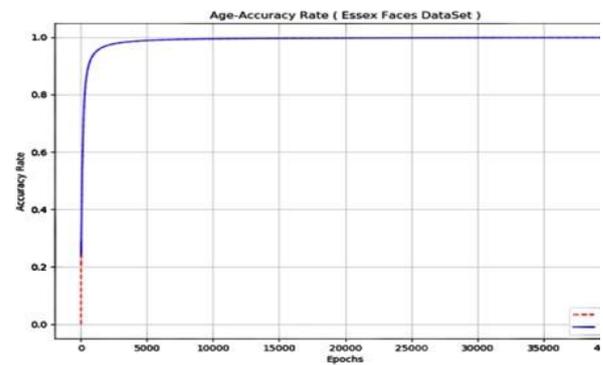


Figure 3: Accuracy rate of age estimation (Essex).

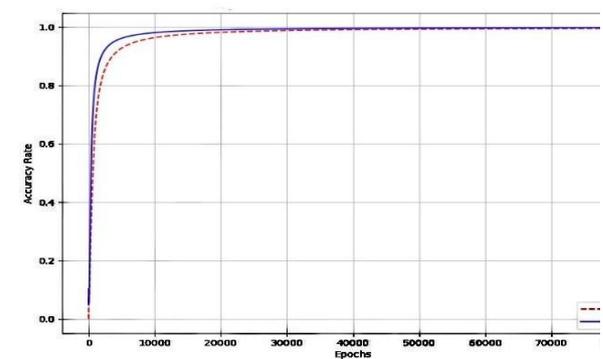


Figure 4: Accuracy of age estimation (Adience).

parameter settings in our experiment listed in Table 1, we plot the loss function and the accuracy function for age then gender separately and jointly.

### 5.2 Results and discussion

In the first, we evaluate our method for classifying a person to the correct age. We train our network to classify face images into six age group classes and report the performance of our classifier on Essex dataset and Adience benchmark.

We also evaluate our method for classifying face images to the correct labels gender. We test the performance on the same database Essex dataset and Adience benchmark. For this task, we train our network for classification of two classes (female and male) and report the result accuracy.

We assess our solution for the age and gender classifications jointly on Essex dataset and Adience benchmark. The purpose is to predict whether a person's gender is within a precise age range.

The proposed work is based on gender and age classification using CNN. We tackled the classification of age group and gender of face images and posed the task as a multiclass classification problem as such train the model with a classification-based loss function as training targets. We perform data preprocessing on the training dataset image by the flatten on a one dimensional vector feature of input image for age and gender. After successfully extracting features, these last will pass to the phase of classification by CNN, the result of the last layer

Parameters	Value
Training iteration	40000 -80000 epochs
Dropout probability (keep dropout)	0.8
Dropout probability (ratsig dropout)	0.2
learning rate	0.00001
Activation function	ReLU
Number of hidden units	512/128
Number of layers (Convolution-Pooling)	4
Classification function for age estimation	Softmax
Classification function for gender	Sigmoid

Table 2: The parameters settings used in our experiments.

Age classes	0-6	8-20	25-32	38-43	48-53	+60
Acc (%)	89.5	93.1	95.3	97.1	99.3	98.2

(a) Essex dataset.

Age classes	0-6	8-20	25-32	38-43	48-53	+60
Acc (%)	83.4	88.6	90.0	95.1	97.2	96.2

(b) Adience benchmark.

Table 3: Gender accuracy per age for the Essex and Adience benchmark datasets.

Method	dataset	Accuracy Gender %	Accuracy Age %
G. Levi and T. Hassner [3]	Essex	N/A	N/A
	Adience benchmark	77.8	45.1
A. Ekmekji [8]	Essex	N/A	N/A
	Adience benchmark	86.8	50.7
A.Olatunbosun et al [18]	Essex	N/A	N/A
	Adience benchmark	96.2	93.8
S.M.Osman et al [20]	Essex	93.93	92.3
	Adience benchmark	N/A	N/A
Our Proposed model	Essex	99.10	95.41
	Adience benchmark	95.6	91.75

Table 4: Age and gender classification: performance comparison of our result with the state-of-the-art works.

of convolutional network is passed to three fully connected layer , finally outputs are given to the fully connected layers to minimize the classification loss by the softmax function on six age classes ( groups 0-6, 8-20, 25-32, 38-43, 48-53, and + 60 years old ), and by the sigmoid function on two gender (female and male) classes .

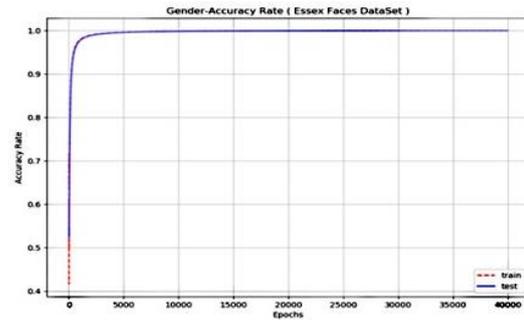


Figure 7: Gender accuracy rate (Essex).

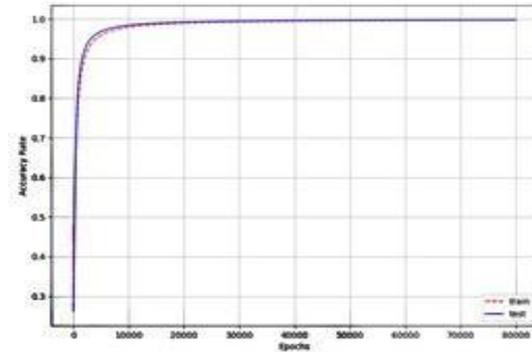


Figure 8: Gender accuracy rate (Adience benchmark).

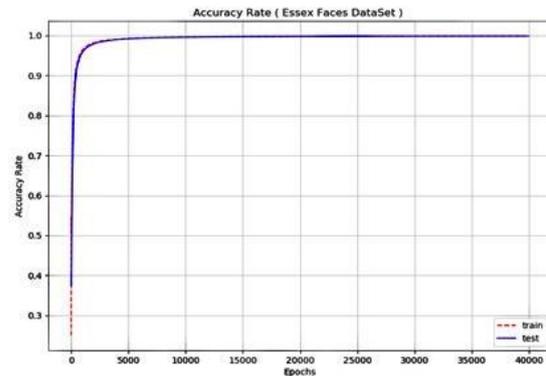


Figure 9: Joint accuracy of age and gender classification (Essex dataset).

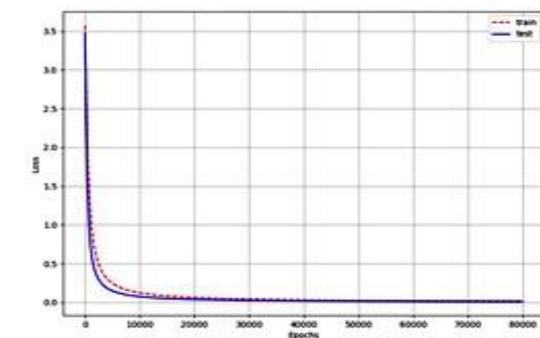


Figure 10: Joint loss of age and gender classification (Adience benchmark).

According to parameter settings in our experiment listed in Table 2, the analysis of Figure (5-6-7-8-9-10) and the result of Table 3 .The proposed network provides significant accuracy improvements of age and gender classification, but requires considerable time for training our network to implement the correct prediction , where

the training with 80000 epoch require a minimum of 36 hours. In Table 3, it is shown that the proposed approach performs very well on adults group age, although he failed to distinguish very young subjects.

This performance is expected as even for humans estimating age the young children is harder than the age of adults. Gender estimation mistakes are also frequently occur for images of babies or very young children where obvious gender attributes are not yet visible.

Some of the state-of-the-art methods for age and gender classification are compared with our model. The results in Table 4 clearly show the improvement accuracy of our method compared to previous work, The proposed models achieve medium accuracy ratio, where their systems make many misclassification, that are due to extremely challenging viewing conditions of some of the Essex and Adience benchmark images. As we note, the best performance was obtained on both Essex face dataset and Adience benchmark. Our approach, therefore, achieves the best results not only on the age estimation but also on gender classification, it outperforms the current state-of-the-art methods.

## 6 Conclusion

In this paper, we evaluated the application of deep neural network for human age and gender prediction using CNN. During this study various design was developed for this task, age and gender classification is one of the key segments of research in the biometric as social applications with the goal that the future forecast and the information disclosure about the particular individual should be possible adequately. The procedure utilizes grouped approaches and calculations whereby the deep learning is likewise the prime in usage designs in the CNN for age and gender classification. A convolutional neural network pipeline for automatic age and gender recognition has been proposed, we provide an extensive data set and benchmark for the study of age estimation and gender classification.

According to the results taken from the proposed age and gender classification methodology from facial images has significantly higher accuracy. The proposed method use facial images in the range of 0 – 68 as it is very difficult to recognize the gender and age of little babies and children by the geometric facial feature variations and also there are not enough facial images of people older than 60 years old to use in the experiment.

In the proposed system we investigate the classification accuracy on Essex dataset and Adience benchmark for age and gender classification, our proposed method achieves the state-of-the-art performance, in both age and gender estimation and significantly outperforming the existing models. For future works, we will consider a pretrained CNN architecture and focus the work to design such a framework for race prediction.

## References

- [1] P. Arumugam, S. Muthukumar, S. Selva Kumar and S. Gayathri, "Human Age Group Prediction from Unknown Facial Image", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 7, pp 722-726, May 2017. <https://doi.org/10.23956/ijarcsse/sv7i5/0103>
- [2] M.K. Benkaddour and A. Bounoua, "Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition", *International Information and Engineering Technology Association IIETA*, 2017. <https://doi.org/10.3166/ts.34.77-91>
- [3] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42, Boston, MA, USA, 2015. <https://doi.org/10.1109/CVPRW.2015.7301352>
- [4] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-cnn for age estimation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 742–751, Honolulu, HI, USA, 2017. <https://doi.org/10.1109/CVPR.2017.86>
- [5] G. Antipov, M. Baccouche, J. Dugelay, "Effective Training of Convolutional Neural Networks for Face-Based Gender and Age Prediction", *Pattern Recognition Volume 72*, 2017. <https://doi.org/10.1016/j.patrec.2015.11.011>
- [6] G. Panis and A. Lanitis, "An Overview of Research Activities in Facial Age Estimation Using the FG-NET Aging Database", *Journal of American History* pp 455-462, 2015. [https://doi.org/10.1007/978-3-319-16181-5\\_56](https://doi.org/10.1007/978-3-319-16181-5_56)
- [7] K. Ricanek and T. Tesafaye, "MORPH: A Longitudinal Image Database of Normal Adult Age-Progression," *7th International Conference on Automatic Face and Gesture Recognition (FG06)*. <https://doi.org/10.1109/fg.2006.78>
- [8] A. Ekmekci, "Convolutional neural networks for age and gender classification," *Technical Report*, Stanford University, 2016.
- [9] E. Eidinger, R. Enbar and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014. <https://doi.org/10.1109/TIFS.2014.2359646>
- [10] K. Fukushima, *Neocognitron: Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. *Biological Cybernetics*, 36, 193-202, 1980. <https://doi.org/10.1007/BF00344251>
- [11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE*, 86, 2278-2324, 1998. <https://doi.org/10.1109/5.726791>
- [12] Essex face database, University of Essex, UK, <http://cswww.essex.ac.uk/mv/allfaces/index.html>
- [13] V. Carletti, A. Greco, G. Percannella, and M. Vento, "Age from Faces in the Deep Learning Revolution,"

- IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 9, pp. 2113–2132, Sep. 2020. <https://doi.org/10.1109/tpami.2019.2910522>
- [14] M. R. Dileep and A. Danti, “Human age and gender prediction based on neural networks and three human age and gender prediction based on neural networks and three sigma control limits,” *Applied Artificial Intelligence*, vol. 32, no. 3, pp. 281–292, 2018. <https://doi.org/10.1155/2020/1289408>
- [15] D. Yi, Z. Lei, and S. Z. Li, “Age Estimation by Multi-scale Convolutional Network,” *Lecture Notes in Computer Science*, pp. 144–158, 2015. [https://doi.org/10.1007/978-3-319-16811-1\\_10](https://doi.org/10.1007/978-3-319-16811-1_10)
- [16] Adience Benchmark, for Gender and Age Classification, <https://talhassner.github.io/home/projects/Adience/Adience-data.html>
- [17] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, “Deep neural network framework and transformed MFCCs for speaker's age and gender classification,” *Knowledge-Based Systems*, Volume 115, Pages 5-14, 2017. <https://doi.org/10.1016/j.knosys.2016.10.008>
- [18] A. Anand, R. D. Labati, A. Genovese, E. Munoz, V. Piuri, and F. Scotti, “Age estimation based on face images and pre-trained convolutional neural networks,” in *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, Honolulu, HI, USA, November 2017. <https://doi.org/10.1109/SSCI.2017.8285381>
- [19] A. Olatunbosun and S. Viriri, “Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces”, *the scientific world journal*, 2020. <https://doi.org/10.1155/2020/1289408>
- [20] D. V. Sang, L. T. Cuong. Effective Deep Multi-source Multi-task Learning Frameworks for Smile Detection, Emotion Recognition and Gender Classification. *Informatica (Slovenia)*, 42, 2018. <https://doi.org/10.31449/inf.v42i3.2301>
- [21] S. M. Osman, N. Noor, and S. Viriri, “Component-Based Gender Identification Using Local Binary Patterns,” *Lecture Notes in Computer Science*, pp. 307–315, 2019. [https://doi.org/10.1007/978-3-030-28377-3\\_25](https://doi.org/10.1007/978-3-030-28377-3_25)

