

Research Article

Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces

Olatunbosun Agbo-Ajala and Serestina Viriri 

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

Correspondence should be addressed to Serestina Viriri; viriris@ukzn.ac.za

Received 15 November 2019; Accepted 11 March 2020

Academic Editor: Susanna Falsaperla

Copyright © 2020 Olatunbosun Agbo-Ajala and Serestina Viriri. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Age and gender predictions of unfiltered faces classify unconstrained real-world facial images into predefined age and gender. Significant improvements have been made in this research area due to its usefulness in intelligent real-world applications. However, the traditional methods on the unfiltered benchmarks show their incompetency to handle large degrees of variations in those unconstrained images. More recently, Convolutional Neural Networks (CNNs) based methods have been extensively used for the classification task due to their excellent performance in facial analysis. In this work, we propose a novel end-to-end CNN approach, to achieve robust age group and gender classification of unfiltered real-world faces. The two-level CNN architecture includes feature extraction and classification itself. The feature extraction extracts feature corresponding to age and gender, while the classification classifies the face images to the correct age group and gender. Particularly, we address the large variations in the unfiltered real-world faces with a robust image preprocessing algorithm that prepares and processes those faces before being fed into the CNN model. Technically, our network is pretrained on an IMDB-WIKI with noisy labels and then fine-tuned on MORPH-II and finally on the training set of the OIU-Adience (original) dataset. The experimental results, when analyzed for classification accuracy on the same OIU-Adience benchmark, show that our model obtains the state-of-the-art performance in both age group and gender classification. It improves over the best-reported results by 16.6% (exact accuracy) and 3.2% (one-off accuracy) for age group classification and also there is an improvement of 3.0% (exact accuracy) for gender classification.

1. Introduction

Facial analysis has gained much recognition in the computer vision community in the recent past [1–4]. Human's face contains features that determine identity, age, gender, emotions, and the ethnicity of people [5, 6]. Among these features, age and gender classification can be especially helpful in several real-world applications including security and video surveillance, electronic customer relationship management, biometrics, electronic vending machines, human-computer interaction, entertainment, cosmetology, and forensic art [7–9]. However, several issues in age and gender classification are still open problems. Age and gender predictions of unfiltered real-life faces are yet to meet the requirements of commercial and real-world applications in spite of the progress computer vision community keeps

making with the continuous improvement of the new techniques that improve the state of the art [10–12].

Over the past years, a lot of methods have been proposed to solve the classifications problem. Many of those methods are handcrafted which perform unsatisfactorily on the age and gender predictions of unconstrained in-the-wild images [2, 13]. These conventional hand-engineered methods relied on the differences in dimensions of facial features and face descriptors [14–20] which do not have the ability to handle the varying degrees of variation observed in these challenging unconstrained imaging conditions. The images in these categories have some variations in appearance, noise, pose, and lighting which may affect the ability of those manually designed computer vision methods to accurately classify the age and gender of the images. Recently, deep learning-based methods [21–28] have shown encouraging

performance in this field especially on the age and gender classification of unfiltered face images. In light of the current works in age and gender classification and encouraging signs of progress in deep learning and CNN, we therefore propose a novel end-to-end deep learning-based classification model that predicts age group and gender of unfiltered in-the-wild facial images. We formulate the age and gender classifications task as a classification problem in which the CNN model learns to predict the age and gender from a face image. Figure 1 displays the overall idea of the proposed model.

The contributions of this work are summarized as follows:

- (1) We propose a model that uses CNN architecture to predict the age group and gender of human's faces from unfiltered real-world environments. The novel CNN approach addresses the age and gender labels as a set of discrete annotations and train the classifiers that predict the human's age group and gender.
- (2) We design a quality and robust image preprocessing algorithm that prepare and preprocess the unfiltered images for the CNN model and this greatly has a very strong impact on the performance accuracy of our age and gender classifiers.
- (3) We demonstrate that pretraining on large-scale datasets allows an effective training of our age and gender CNN model which enable the classifiers to generalize on the test images and then avoid overfitting.
- (4) Finally, OIU-Adience benchmark is used to evaluate the performance of our novel CNN model, and despite the very challenging nature of the images in the dataset, our approach produces significant improvements in age group and gender classification accuracy over the state-of-the-art methods; the result can satisfy the requirements of several real-world applications.

The rest of the paper is arranged as follows: in Section 2, we present related works in age and gender classification; in Section 3, we describe the proposed method for the classification task; in Section 4, we detail the experimental analysis and the results when evaluated on OIU-Adience dataset benchmark; and we summarize our contributions and conclusions with future works in Section 5.

2. Related Works

In this section, we briefly review the age and gender classification literature and describe both the early methods and those that are most related to our proposed method, focussing on age and gender classification of face images from unconstrained real-world environments. Almost all of the early methods in age and gender classifications were handcrafted, focussing on manually engineering the facial features from the face and mainly provides a study on constrained images that were taken from controlled imaging

conditions. To mention a few, in 1999, Kwon and Lobo [29] developed the very first method for age estimation focussing on geometric features of the face that determine the ratios among different dimensions of facial features. These geometric features separate babies from adult successfully but are incapable of distinguishing between young adult and senior adult. Hence, in 2004, Lanitis et al. [30] proposed an Active Appearance Model (AAM) based method that included both the geometric and texture features, for the estimation task. This method is not suitable for the unconstrained imaging conditions attributed to real-world face images which have different degrees of variations in illumination, expression, poses, and so forth. From 2007, most of the approaches also employed manually designed features for the estimation task: Gabor [14], Spatially Flexible Patches (SFP) [17], Local Binary Patterns (LBP) [31, 32], and Biologically Inspired Features (BIF) [33]. In recent years, classification and regression methods are employed to classify the age and gender of facial images using those features. Classification methods in [12, 34–36] used Support Vector Machine (SVM) based methods for age and gender classification. Linear regression [20, 37], Support Vector Regression (SVR) [38], Canonical Correlation Analysis (CCA) [39], and Partial Least Squares (PLS) [40] are the common regression methods for age and gender predictions. Dileep and Danti [41] also proposed an approach that used feed-forward propagation neural networks and 3-sigma control limits approach to classify people's age into children, middle-aged adults, and old-aged adults. However, all of these methods are only suitable and effective on constrained imaging conditions; they cannot handle the unconstrained nature of the real-world images and, therefore, cannot be relied on to achieve respectable performance on the in-the-wild images which are common in practical applications [12].

More recently, an expanding number of researchers start to use CNN for age and gender classification. It can classify the age and gender of unfiltered face images relying on its good feature extraction technique [23, 26–28, 42–44]. Availability of sufficiently large data for training and high-end computer machines also help in the adoption of the deep CNN methods for the classification task. CNN model can learn compact and discriminative facial features, especially when the volume of training images is sufficiently large, to obtain the relevant information needed for the two classifications. For example, in 2015, Levi et al. [13] proposed a CNN based model, comprising of five layers, three convolutional and two fully connected layers, to predict the age of real-world face images. The model included centre-crop and oversampling method, to handle the small misalignment in unconstrained images. Yi et al. [45], in their paper, applied an end-to-end multitask CNN system that learns a deeper structure and the parameters needed, to solve the age, gender, and ethnicity classification task. In [46], the authors investigated a pretrained deep VGG-Face CNN approach, for automatic age estimation from real-world face images. The CNN based model consists of eleven layers, including eight convolutional and three fully connected layers. The authors in [2] also proposed a novel CNN based method, for

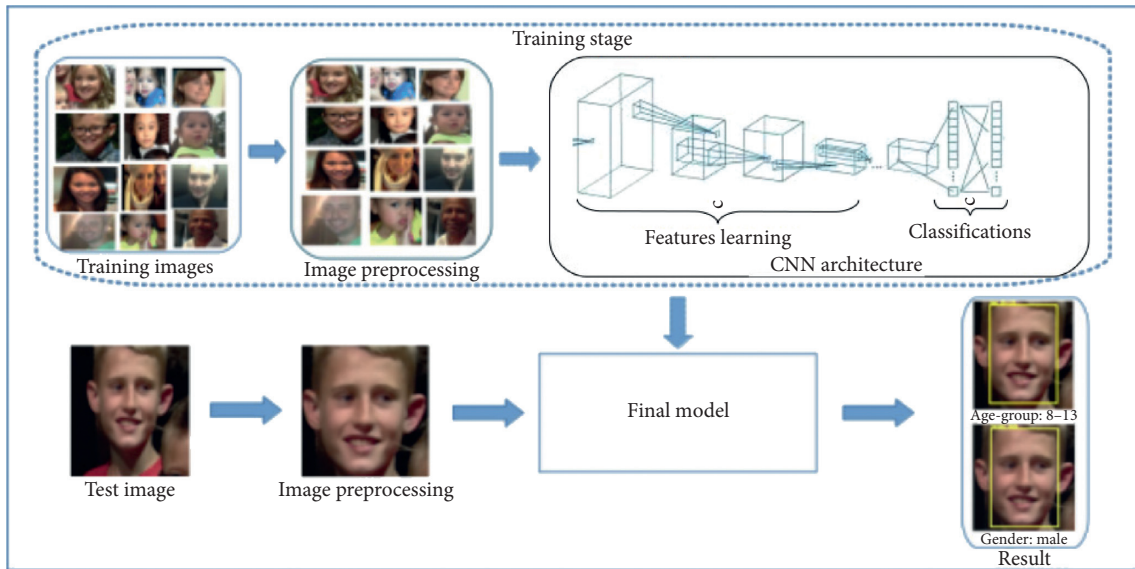


FIGURE 1: The pipeline of our framework for age group and gender.

age group and gender estimation: Residual Networks of Residual Networks (RoR). The model includes an RoR architecture, which was pretrained on gender and weighted loss layer and then on ImageNet dataset, and finally it was fine-tuned on IMDb-WIKI-101 dataset. Ranjan et al. in [47] presented a model that simultaneously solves a set of face analysis tasks, using a single CNN. The end-to-end solution is a novel multitask learning CNN framework, which shares the parameters from lower layers of CNN among all the tasks for gender recognition, age estimation, etc. In [9], the authors proposed a CNN solution for age estimation, from a single face image. The CNN based solution includes a robust face alignment phase that prepares and preprocesses the face images before being fed to the designed model. The authors also collected large-scale face images, with age and gender label: IMDb-WIKI dataset. In 2018, Liu et al. [48] developed a CNN based model that employed a multiclass focal loss function. The age estimation model was validated on Adience benchmark for performance accuracy, and it achieved a comparable result with state-of-the-art methods. Also in [49], Duan et al. introduced a hybrid CNN structure for age and gender classification. The model includes a CNN and Extreme Learning Machine (ELM). The CNN extracts the features from the input images while ELM classifies the intermediate results. In [50], the authors proposed a robust estimations solution (CNN2ELM) that also includes a CNN and ELM. The model, an improvement of the work in [49], is three CNN based solutions for age, gender, and race classification from face images. The authors in [51] also proposed a novel method based on “attention long short-term memory (LSTM) network” for age estimation in-the-wild. The method was evaluated on Adience, MORPH-II, FG-NET, LAP15, and LAP16 datasets for performance evaluation. Also in [52], the authors studied an age group-n encoding CNN based model: AGen. The model explores the relationship between the real age and its adjacent ages, by grouping adjacent ages into the same group.

Unfortunately, some of these methods mentioned above have been verified effectively on constrained imaging conditions; few studied the unconstrained imaging conditions. However, it is still a challenging problem classifying unconstrained faces with large variations in illumination, viewpoint, nonfrontal, etc. There is a need for a suitable and robust model that can improve the state-of-the-art methods for its applicability in intelligent and real-world applications. Here, we address those issues by designing a robust image preprocessing algorithm, pretrain the model on large-scale facial aging benchmarks with noisy age and gender labels, and also regularize the CNN parameters with our novel CNN framework.

3. Proposed Method

This section presents the proposed deeply learned classifiers for age group and gender classification of unfiltered real-life face images, an extension of our conference paper in [53].

The approach as presented in Algorithm 1 requires image preprocessing (face detection, landmark detection, and face alignment) stage that preprocess and prepare the face images before they are input into the proposed network. Therefore, our solution is divided into three major steps: image preprocessing, features learning, and classification itself.

3.1. Image Preprocessing. Intelligent age and gender classifiers tackle the classification task under unfiltered real-world settings. Most of those face images are not aligned and nonfrontal and also with different degrees of variations in pose, appearance, lighting, and background conditions. Therefore, those in-the-wild face images need first to be detected, then aligned, and, finally, used as input for the classifiers. The image preprocessing phase as shown in Figure 2 is explained in more details below.

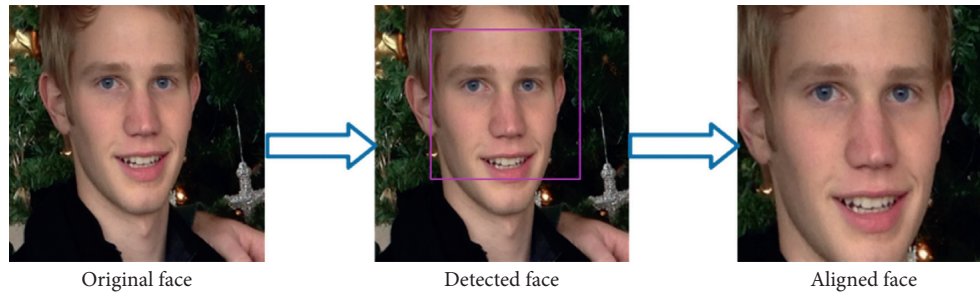


FIGURE 2: The image preprocessing phase.

3.2. Face Detection. The first stage of image preprocessing is face detection. The face detection phase locates the face in an input image. In this work, we employ an open-source face detector: Head Hunter described in [54]. In order to detect a face, all the input images are rotated in the range of -90° to 90° angles and with the step of 5° . After that, the detector selects the input image with the best output of the face detector and in a case where the face is not detected in all the modifications of the input image, the original input image is upscaled and face detection algorithm is repeated until a face is detected. The upscaling helps in detecting faces in all the input images.

3.3. Landmark Detection and Face Alignment. After face detection, is the facial landmark detection and face alignment phase, where we employ the state-of-the-art solution in [55]. This image preprocessing solution is an open-source multiview facial landmark detection algorithm that uses five landmark detection models, including a frontal model, two half-profile models, and two full profile models. All of these five models are trained to work on one of the corresponding facial poses. The face alignment phase, on the other hand, entails running of all the five facial landmark models, on the detected faces. An affine transformation is then performed on the model, with the highest confidence score, to the predefined optimal settings of those landmarks.

3.4. CNN Architecture. Our CNN architecture is a novel six-layer network, comprising four convolutional and two fully connected layers. The CNN design is an end-to-end sequential deep learning architecture, including feature extraction and classification phases. The feature extraction phase has four convolutional layers, with the corresponding parameters, including the number of filters, the kernel size of each filter, and the stride. It contains the convolutional layer, activation layer (rectified linear unit (ReLU)), batch normalization (instead of the conventional Local Response Normalization), max-pooling layer, and a dropout [56] (all the convolutional layers have a fixed dropout of 25%). The classification stage, on the other hand, contains two fully connected layers, that handle the classification phase of the model. The first fully connected layer contains 512 neurons, followed by a ReLU, then batch normalization, and, finally, a dropout layer at a dropout ratio of 0.5. The second and the last fully connected layer output 512 features which are

densely mapped to 8 or 2 neurons for classification tasks. The details of the CNN structure are presented in Table 1.

We model age group and gender classification task as an end-to-end deep classification problem; therefore, a softmax with cross entropy loss function is adopted to obtain a probability for each age group and gender class. We present the design of some parameter of the CNN architecture in the section below.

Softmax classifier gives you probabilities for each class label and its function is defined as follows:

$$f_j(s) = \frac{e^{s_j}}{\sum_k e^{s_k}}, \quad (1)$$

where we are using the notation f_j to mean the j th element of the vector of class scores f that takes a vector of arbitrary real-valued scores in s .

A cross entropy loss is used for training the multiclass and binary classifications of age and gender classifiers.

We define the cross entropy for binary classification as follows:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N [x_i \log(p(\hat{x}_i)) + (1 - x_i) \log(1 - p\hat{x}_i)], \quad (2)$$

where x is the binary class label, 1 if it is the correct class and 0 otherwise, and $p(x)$ is the predicted probability of the point being green for all N points.

For multiclass, cross entropy is defined as follows:

$$H_{x'}(x) := - \sum_i x'_i \log(x_i), \quad (3)$$

where x_i is the predicted probability value for class i and x'_i is the true probability for that class.

3.5. Training Details. In this section, we describe the training details for age group and gender classifiers on IMDB, MORPH-II, and OIU-Adience datasets benchmark. The age group classifier will be responsible for predicting the age groups of unfiltered human's face images into eight different classes, while gender classifier will classify those face images into two gender classes.

For all our experiments, we initialized and trained our CNN model from scratch, using the images and the labels of IMDB and MORPH-II datasets benchmark. We primarily pretrained the novel CNN architecture on the IMDB-WIKI

Input: the training face images $\{x_i, y_i, g_i\}_{i=1, \dots, m}$ and the test images $\{\bar{x}_j, \bar{y}_j, \bar{g}_j\}$.
Output: the predictions for the input test images $\{\bar{y}_j, \bar{g}_j\}_{j=1, \dots, n}$.

- (1) Perform preprocessing for the training images to obtain aligned images.
- (2) Train the novel CNN with the training images, $\mathbf{X} = \{x_i\}_{i=1, \dots, m}$, to obtain the age and gender classifiers $h()$ and $f()$, respectively.
- (3) **For** $j = 1$ to n
- (4) Perform preprocessing for the test images to obtain aligned images.
- (5) Input the aligned face images into the trained CNN classifier.
- (6) *If* the CNN is age classifier, $h()$
- (7) **return** $\bar{y}_j \leftarrow h(\bar{x}_j)$
- (8) *Else*
- (9) **return** $\bar{g}_j \leftarrow f(\bar{x}_j)$
- (10) **End For**

ALGORITHM 1: Algorithm for the age and gender classifiers.

TABLE 1: The details of CNN architecture for the proposed method.

Layer type	Output size	Filter size/stride
Input image	$227 \times 227 \times 3$	—
CONV1	$56 \times 56 \times 96$	$7 \times 7/4 \times 4$
ACT	$56 \times 56 \times 96$	—
BN	$56 \times 56 \times 96$	—
Maxpool	$28 \times 28 \times 96$	$3 \times 3/2 \times 2$
Dropout	$28 \times 28 \times 96$	—
CONV2	$28 \times 28 \times 256$	5×5
ACT	$28 \times 28 \times 256$	—
BN	$28 \times 28 \times 256$	—
Maxpool	$14 \times 14 \times 256$	3×3
Dropout	$14 \times 14 \times 256$	—
CONV3	$14 \times 14 \times 384$	3×3
ACT	$14 \times 14 \times 384$	—
BN	$14 \times 14 \times 384$	—
Maxpool	$7 \times 7 \times 384$	3×3
Dropout	$7 \times 7 \times 384$	—
CONV4	$7 \times 7 \times 384$	3×3
ACT	$7 \times 7 \times 256$	—
BN	$7 \times 7 \times 256$	—
Maxpool	$1 \times 1 \times 256$	3×3
Dropout	$1 \times 1 \times 256$	—
FC1	512	—
ACT	512	—
BN	512	—
Dropout	512	—
FC2	2 or 8	—

unfiltered facial aging dataset whose images are obtained directly from the website with some degree of variability and then fine-tuned the CNN on the images from the MORPH-II dataset, to avoid overfitting and also to adapt the CNN model to face image contents of the task to perform. Finally, we tuned the network on the training part of the actual dataset (OIU-Adience) on which we evaluated. The fine-tuning allows the CNN to pick up the distribution, the particularities, and the bias of each dataset, hence improving the performance. For each of the datasets of IMDB-WIKI and MORPH-II, we split it into two: 90% for training and 10% for validation. On the original OIU-Adience dataset,

training and testing for both age and gender classification are performed using the standard 5-fold cross-validation procedure that is defined in recent literature.

3.6. Age Group Classification. To train the age group CNN based classifier to predict unfiltered face images into the correct age group and after series of empirical experiments, we set the initial learning rate to be 0.0001 to allow model train longer and then use an $L2$ weight decay of 0.0005. To make our model able to generalize and predict correctly, we apply Adam optimizer [57] to update network weights during training.

We present the updated equations (see equations (4)–(6)) for Adam optimization algorithm as follows:

For each parameter w^j

$$m_t = \gamma_1 * m_{t-1} + (1 - \gamma_1) * g_t, \quad (4)$$

$$v_t = \gamma_2 * v_{t-1} + (1 - \gamma_2) * g_t^2,$$

$$\Delta w_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} * g_t, \quad (5)$$

$$w_{t+1} = w_t + \Delta w_t, \quad (6)$$

where η is the learning rate, g_t is the gradient at time t along w^j , m_t and v_t are the exponential average of squares of gradients along w_j , and γ_1 and γ_2 are the hyperparameters.

3.7. Gender Classification. Training the gender classifier, we set the learning rate to an initial value of 0.01, L_2 weight decay to 0.0005, and momentum term to 0.9. In this case, we implement a faster and popular optimization algorithm: Stochastic Gradient Descent (SGD), to update all the training models. SGD produces an update to the parameter of each training input $x^{(i)}$ and label $y^{(i)}$.

SGD is defined as follows:

$$\gamma = \gamma - \eta \cdot \Delta_\gamma J(\gamma; x^{(i)}; y^{(i)}), \quad (7)$$

where η is defined as the learning rate and $\Delta_\gamma J$ is the gradient of the loss term with respect to the weight vector γ .

A summary of the training details is presented in Table 2.

4. Proposed Algorithm

The proposed algorithm for CNN model is given in Algorithm 1.

5. Experiments

In this section, we first introduce the datasets with a description of their specifications and present the experimental analysis and results of the experiments when evaluated for age group and gender classifications accuracy.

5.1. Datasets. For the age group and gender classification, we evaluate the proposed method on OIU-Adience [12] dataset. IMDB-WIKI [2] and MORPH-II [58] datasets are also employed to pretrain our network when evaluating the classifiers on OIU-Adience dataset. A brief summary of those datasets is given in Table 3, with the size of each dataset, age-range information, and the number of subjects. Figure 3 also presents some sample images of each dataset. Here, we take a brief introduction of those datasets and their specifications.

5.2. IMDB-WIKI [2]. IMDB-WIKI is the largest publicly available dataset for age estimation of people in the wild, containing more than half a million images with accurate age labels between 0 and 100 years. The images were crawled

from IMDB and Wikipedia. IMDB contains 460,723 images of 20,284 celebrities and Wikipedia contains 62,328 images obtained directly from the website, as such the dataset contains many low-quality images, such as human comic images, sketch images, severe facial mask, full body images, multiperson images, and blank images.

5.3. MORPH-II [58]. MORPH-II is the mostly used dataset for real age estimation. It is a publicly available facial aging benchmark with about 55,000 facial images from more than 13,000 subjects. MORPH-II comprises 46,645 images of males and 8487 images of females with an age range from 16 to 77 years.

5.4. OIU-Adience [12]. OIU-Adience is a collection of face images from ideal real-life and unconstrained environments. It reflects all the features that are expected from an image collected from challenging real-world scenarios. They are face images that were uploaded to Flickr website from smartphone without any filtering. Adience images, therefore, display a high-level of variations in noise, pose, and appearance, among others. The entire collection of OIU-Adience dataset is about 26,580 face images of 2,284 subjects and with an age group label of eight comprising 0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, and 60+. The distribution of the face images for age group and gender class labels in OIU-Adience benchmark is presented in Figure 4.

5.5. Experimental Results and Analysis. In this section, we present the results of our experiments on the validation set of the OIU-Adience dataset describing the impacts of our design choices on the age group and gender classification performance. We employ two different evaluating metrics common in the literature, to assess the performance of the two classifiers: exact accuracy and one-off accuracy.

5.6. Exact Accuracy. Exact accuracy [13] calculates the exact age group and gender results. It measures the percentage of face images that were classified into correct age group and gender, which is the ratio of the accurate predictions to the total number of the ground-truth labels. Mathematically, the metric is described as follows:

$$\text{Exact accuracy} = \frac{\text{No. of accurate prediction}}{\text{Total no. of prediction}}. \quad (8)$$

5.7. One-Off Accuracy. One-off accuracy [12] is off by one adjacent age group and measures whether the ground-truth class label matches the predicted class label or if the ground-truth label exists in the two adjacent bins.

We present the result for age group classification using the two criteria while the result for gender classification will be presented using exact accuracy metric.

5.8. Age Group Classification. We evaluate our method for classifying a person to the correct age group. We train our

TABLE 2: Training details with OIU-Adience.

Classifier	Optimizer	No. of epochs	Initial learning rate	Momentum term	L2 weight decay
Age group	Adam	150	0.0001	—	0.0005
Gender	SGD	140	0.01	0.9	0.0005

TABLE 3: The details of the datasets used in our experiments.

Dataset	Dataset size	No. of subjects	Age type	Age range
IMDb-WIKI [2]	523,051	20,284	Real age	0–100
MORPH-II [58]	55,134	13,618	Real age	16–77
OIU-Adience [12]	26,580	2284	Age group	0–60+

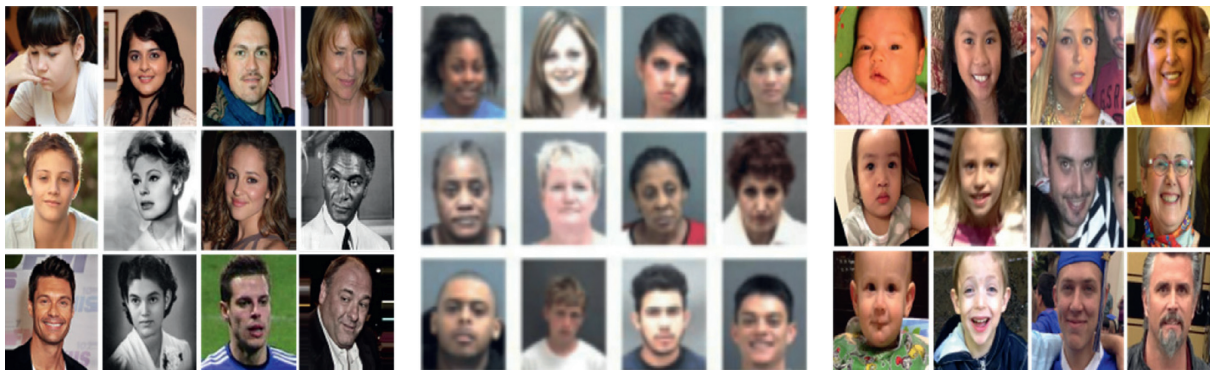


FIGURE 3: Sample face images from IMDb-WIKI, MORPH-II, and OIU-Adience datasets.

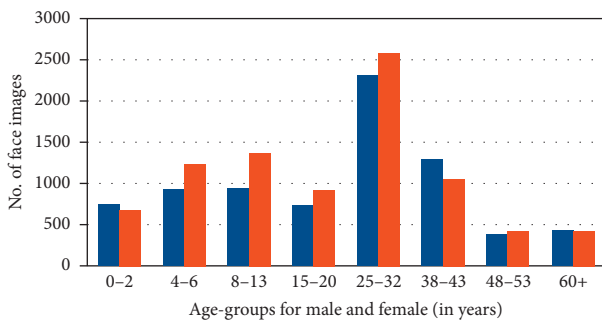


FIGURE 4: The number of image samples for age group and gender in OIU-Adience dataset.

network to classify face images into eight age group classes and report the performance of our classifier on OIU-Adience dataset, a standard dataset benchmark for the existing methods for age group and gender classification. Our model, when evaluated on the dataset, obtains an exact accuracy of 83.1% and one-off accuracy of 93.8%. This improves over the best-reported state-of-the-art result for exact accuracy in Duan et al. [50] by 16.6%, and an improvement of 3.2% on one-off result was reported in Qawaqneh et al. [46]. The graphs in Figures 5(a) and 5(b) present the exact and one-off results for the age group classification, on the validation set of OIU-Adience dataset.

5.9. Gender Classification. We also evaluate our method for classifying a person to the correct gender. We assess the performance on the same Adience dataset consisting of

labels for gender. For this task, we train our network for classification of two classes and report the result on exact accuracy, with pretraining on the three datasets. As presented in Figure 5(c), we achieve an accuracy of 96.2% compared to the previous state-of-the-art of 93.2% in Zhang et al. [2] showing an improvement of 3.0%. Our approach, therefore, achieves the best results not only on the age group estimation but also on gender classification; it outperforms the current state-of-the-art methods. Tables 4 and 5 present the classification results of the best configuration of our method with that of the state-of-the-art methods on OIU-Adience benchmark dataset. All the reported results have been computed using the same 5-fold cross-validation protocol. We also present the loss function results for the two classifications in Figures 6(a) and 6(b).

In the first line of Table 6, we present the results for the age group and gender classifications, when pretrained on IMDb-WIKI dataset. These results are to be compared with the results in line 2 which represent the performance of our model fine-tuned on the training set of another large facial aging dataset (MORPH-II). A huge improvement in the results from line 1 shows clearly the importance of fine-tuning our CNN model on huge training images with age and gender annotations, to learn from the representation power of the CNN and pick up the bias and peculiarities of the dataset. Line 3 of Table 6, highlights the impact of the quality of a robust image preprocessing algorithm, on the classification accuracy. The data augmentation during the fine-tuning on the original (OIU-Adience) dataset (line 4 of Table 6) proved to be very efficient as well, with an improvement of the results from line 3 (fine-tuning without

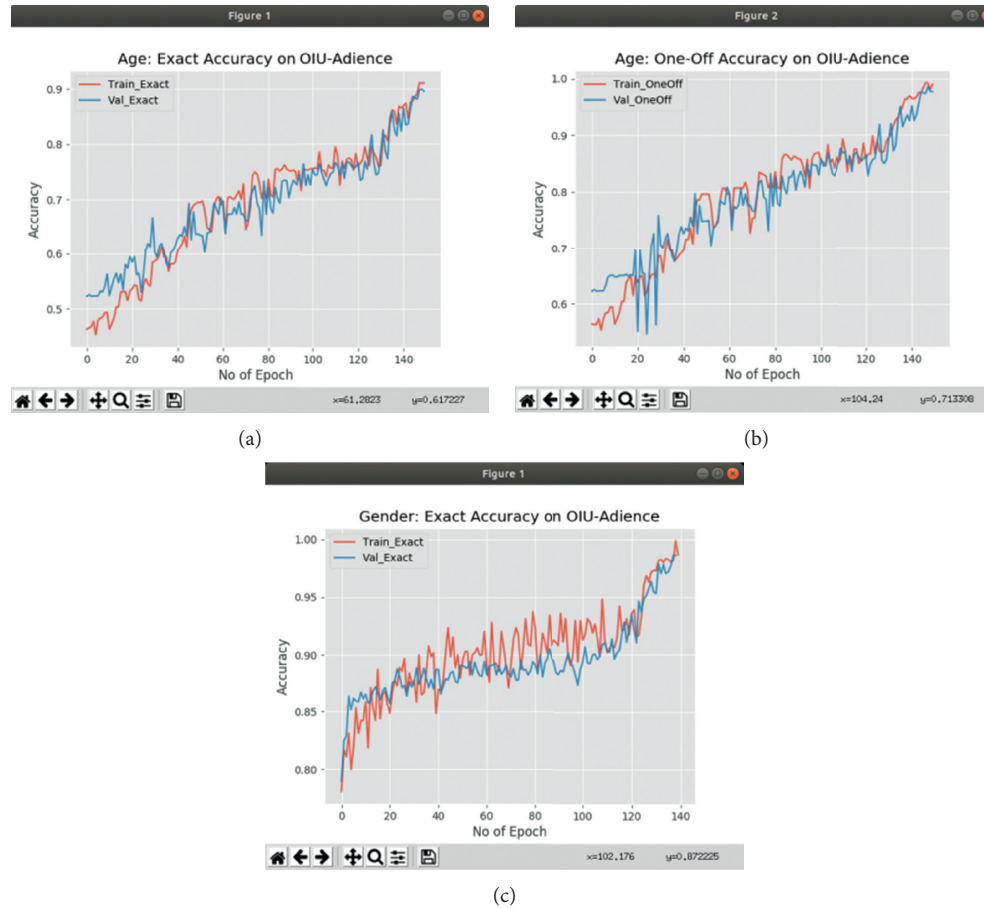


FIGURE 5: Graphs of accuracy results for the age group and gender classification. (a) Exact accuracy for age classifier. (b) One-off accuracy for age classifier. (c) Exact accuracy for gender classifier.

TABLE 4: Age group classification: performance comparison of our result with the state-of-the-art works on OIU-Adience dataset.

Methods	Exact accuracy (%)	One-off (%)
[12]	45.1	79.5
[13]	50.7	84.7
[59]	58.5	—
[46]	59.9	90.6
[49]	52.3	—
[48]	54.0	88.2
[50]	66.5	—
<i>Proposed</i>	83.1	93.8

TABLE 5: Gender classification: performance comparison of our result with the state-of-the-art works on OIU-Adience dataset.

Methods	Exact accuracy (%)
[13]	86.8
[2]	93.2
[49]	88.2
<i>Proposed</i>	96.2

data augmentation). Finally, the last line of Table 6 proves the importance of utilizing the Batch Normalization (BN) rather than the conventional and deprecated Local Response

Normalization (LRN); this improved the classification accuracy of our model.

Some samples of validation classification results are shown in Figure 7, where we present the age group and gender predictions of some of the face images from the OIU-Adience dataset using the newly designed model. In many cases, our solution is able to correctly predict the age group and gender of faces. Undoubtedly, our proposed model did result in some cases of age group and gender misclassification. Some of those misclassification results are presented in Figure 8. The classification failure could be as a result of the extremely challenging viewing conditions of the OIU-Adience images including low

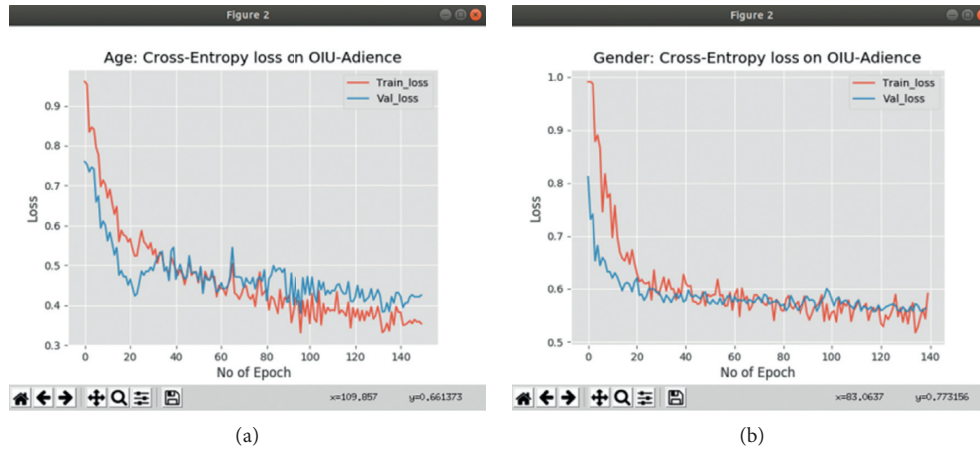


FIGURE 6: Graphs of cross entropy loss results for the age group and gender classification. (a) Cross entropy loss for age classifier. (b) Cross entropy loss for gender classifier.

TABLE 6: The details of experimental results of our model on OIU-Adience dataset.

Pretraining on IMDb-WIKI	Fine-tuning on MORPH-II	Image preprocessing	Data augmentation	Batch normalization	Exact acc. (age)	One-off (age)	Exact acc. (gender)
Yes	No	No	No	No	71.2	84.8	91.3
Yes	Yes	No	No	No	76.1	88.3	93.8
Yes	Yes	Yes	No	No	79.3	90.6	94.5
Yes	Yes	Yes	Yes	No	81.2	91.8	95.9
Yes	Yes	Yes	Yes	Yes	83.1	93.8	96.2

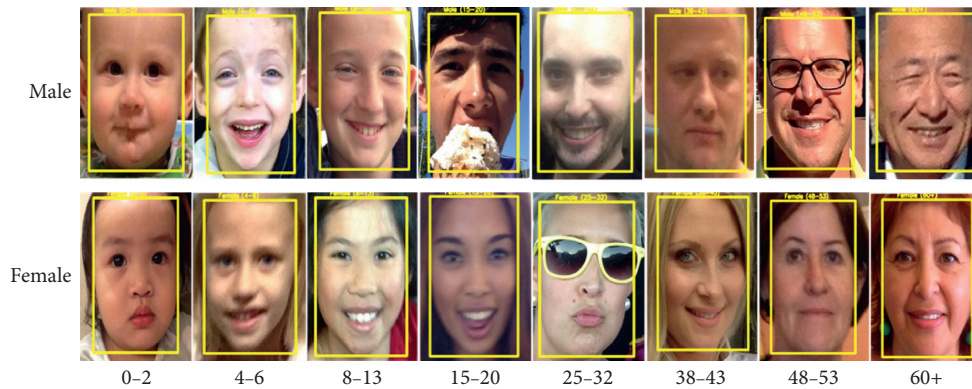


FIGURE 7: Samples of face images with correct age group and gender classifications.

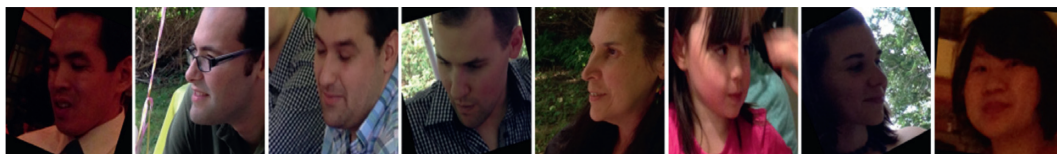


FIGURE 8: Samples of face images with misclassified age group or gender.

resolution, lighting conditions, and heavy makeup, hence hindering our image preprocessing algorithm in either correctly detecting or aligning the face for the classification process.

6. Conclusion and Future Works

We tackled the classification of age group and gender of unfiltered real-world face images. We posed the task as a

multiclass classification problem and, as such, train the model with a classification-based loss function as training targets. Our proposed model is originally pretrained on age and gender labelled large-scale IMDB-WIKI dataset, whose images are obtained directly from the website with some degree of variability and then fine-tuned on MORPH-II, another large-scale facial aging dataset with age and gender annotations. Finally, we use the original dataset (OIU-Adience benchmark of unfiltered faces for age and gender classification) to fine-tune this model. The robust image preprocessing algorithm, handles some of the variability observed in typical unfiltered real-world faces, and this confirms the model applicability for age group and gender classification in-the-wild. Finally, we investigate the classification accuracy on OIU-Adience dataset for age and gender; our proposed method achieves the state-of-the-art performance, in both age group and gender classification, significantly outperforming the existing models. For future works, we will consider a deeper CNN architecture and a more robust image processing algorithm for exact age estimation. Also, the apparent age estimation of human's face will be interesting research to investigate in the future.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] 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 *Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Biometrics Wild, Bwild, Washington, DC, USA, pp. 87–94, June 2017.
- [2] K. Zhang, C. Gao, L. Guo et al., "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492–22503, 2017.
- [3] A. Kuehlkamp, "Age estimation from face images," in *Proceedings of the 6th IAPR International Conference on Biometrics (ICB)*, pp. 1–10, Madrid, Spain, June 2013.
- [4] V. Carletti, A. S. Greco, G. Percannella, M. Vento, and I. Fellow, "Age from faces in the deep learning revolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2019.
- [5] B. Bin Gao, H. Y. Zhou, J. Wu, and X. Geng, "Age estimation using expectation of label distribution learning," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 712–718, Stockholm, Sweden, July 2018.
- [6] R. C. Malli, M. Aygun, and H. K. Ekenel, "Apparent age estimation using ensemble of deep learning models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 714–721, Las Vegas, NV, USA, June 2016.
- [7] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 801–809, Las Vegas, NV, USA, June 2016.
- [8] G. Antipov, M. Baccouche, S. A. Berrani, and J. L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15–26, 2017.
- [9] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, no. 2–4, pp. 144–157, 2018.
- [10] H. Han and A. K. Jain, "Age, gender and race estimation from unconstrained face images," MSU Technical Report, MSU-CSE-14-5, Michigan State University, East Lansing, MI, USA, 2014.
- [11] J. Huang, B. Li, J. Zhu, and J. Chen, "Age classification with deep learning face representation," *Multimedia Tools and Applications*, vol. 76, no. 19, pp. 20231–20247, 2017.
- [12] 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.
- [13] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42, Boston, MA, USA, June 2015.
- [14] F. Gao and H. Ai, "Face age classification on consumer images with gabor feature and fuzzy LDA method," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5558, pp. 132–141, LNCS, Springer Science+Business Media, Berlin, Germany, 2009.
- [15] A. Günay and V. Vasif, "Age estimation based on AAM and 2D-DCT features of facial images," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 2, pp. 113–119, 2015.
- [16] A. Günay and V. Vasif, "Facial age estimation based on decision level fusion of AAM, LBP and gabor features," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 8, pp. 19–26, 2015.
- [17] S. Yan, M. Liu, and T. S. Huang, "Extracting age information from local spatially flexible patches," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 737–740, Las Vegas, NV, USA, March 2008.
- [18] M. Y. El Dib and M. El-saban, *Human Age Estimation Using Enhanced Bio-Inspired Features (EBIF)*, Faculty of Computers and Information, Cairo University, Cairo, Egypt, 2010.
- [19] L. Cai, L. Huang, and C. Liu, "Age estimation based on improved discriminative Gaussian process latent variable model," *Multimedia Tools and Applications*, vol. 75, no. 19, pp. 11977–11994, 2016.
- [20] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 578–584, 2008.
- [21] R. Ranjan, S. Zhou, J. C. Chen et al., "Unconstrained age estimation with deep convolutional neural networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 351–359, Santiago, Chile, December 2015.
- [22] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, "First steps towards an intelligent laser welding

- architecture using deep neural networks and reinforcement learning,” *Procedia Technology*, vol. 15, pp. 474–483, 2014.
- [23] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.
- [24] S. Karen and Z. Andrew, “Very deep convolutional networks for large-scale image recognition,” *Information and Software Technology*, vol. 51, no. 4, pp. 769–784, 2015.
- [25] X. Wang, R. Guo, and C. Kambhamettu, “Deeply-learned feature for age estimation,” in *Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 534–541, Waikoloa, HI, USA, January 2015.
- [26] 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 (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [27] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Medical image computing and computer-assisted Intervention–MICCAI,” *ICLR*, vol. 8150, pp. 1–13, 2017.
- [28] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, Honolulu, HI, USA, 2017.
- [29] Y. H. Kwon and N. Da Vitoria Lobo, “Age classification from facial images,” *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.
- [30] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing different classifiers for automatic age estimation,” *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 621–628, 2004.
- [31] A. Günay and V. V. Nablyev, “Automatic age classification with LBP,” in *Proceedings of the 2008 23rd International Symposium on Computer and Information Sciences*, pp. 6–9, Istanbul, Turkey, 2008.
- [32] S. E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, and A. Benlamoudi, “Facial age estimation using BSIF and LBP,” 2014, <https://arxiv.org/abs/1601.01876>.
- [33] G. Guo, G. Mu, Y. Fu, and T. S. Huang, “Human age estimation using bio-inspired features,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1589–1592, Miami, FL, USA, 2009.
- [34] M. A. Beheshti-nia and Z. Mousavi, “A new classification method based on pairwise support vector machine (SVM) for facial age estimation,” *Journal of Industrial and Systems Engineering*, vol. 10, no. 1, pp. 91–107, 2017.
- [35] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, “Age estimation using a hierarchical classifier based on global and local facial features,” *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.
- [36] S. E. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, “Facial age estimation and gender classification using multi level local phase quantization,” in *Proceedings of the 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, Tlemcen, Algeria, May 2015.
- [37] A. Demontis, B. Biggio, G. Fumera, and F. Roli, “Super-sparse regression for fast age estimation from faces at test time,” *Image Analysis and Processing—ICIAP*, Springer, Berlin, Germany, 2015.
- [38] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, “Image-based human age estimation by manifold learning and locally adjusted robust regression,” *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.
- [39] Y. Fu and G. Guo, “Age synthesis and estimation via faces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.
- [40] G. Guo and G. Mu, “Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression,” in *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 657–664, Colorado Springs, CO, USA, June 2011.
- [41] 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.
- [42] M. Lin, Q. Chen, and S. Yan, “Network in network,” 2013, <https://arxiv.org/abs/1312.4400>.
- [43] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [44] C. Belver, I. Arganda-Carreras, and F. Dornaika, “Evaluating age estimation using deep convolutional neural nets,” *Electronic Imaging*, vol. 2017, no. 17, pp. 100–105, 2017.
- [45] D. Yi, Z. Lei, and S. Z. Li, “Age estimation by multi-scale convolutional network,” *Computer Vision—ACCV 2014*, pp. 144–158, Springer, Berlin, Germany, 2015.
- [46] Z. Qawaqneh, A. A. Mallouh, and B. D. Barkana, “Deep convolutional neural network for age estimation based on VGG-face model,” 2017, <https://arxiv.org/abs/1709.01664>.
- [47] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, “An all-in-one convolutional neural network for face analysis,” in *Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Biometrics Wild, Bwild, Washington, DC, USA, pp. 17–24, June 2017.
- [48] W. Liu, L. Chen, and Y. Chen, “Age classification using convolutional neural networks with the multi-class focal loss,” *IOP Conference Series: Materials Science and Engineering*, vol. 428, no. 1, 2018.
- [49] 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, 2018.
- [50] M. Duan, K. Li, and K. Li, “An ensemble CNN2ELM for age estimation,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 758–772, 2018.
- [51] K. Zhang, N. Liu, X. Yuan, S. Member, X. Guo, C. Cao et al., “Fine-grained age estimation in the wild with attention LSTM networks,” 2019, <https://arxiv.org/abs/1805.10445>.
- [52] Z. Tan, J. Wan, Z. Lei, R. Zhi, G. Guo, and S. Z. Li, “Efficient group-n encoding and decoding for facial age estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2610–2623, 2018.
- [53] O. Agbo-Ajala and S. Viriri, “Age group and gender classification of unconstrained faces,” in *Advances in Visual Computing*, pp. 418–429, Springer, Berlin, Germany, 2019.
- [54] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool, “Face detection without bells and whistles,” *Computer Vision—ECCV 2014*, Springer, vol. 8692, pp. 720–735, Berlin, Germany, 2014.
- [55] M. Uříčář, V. Franc, D. Thomas, A. Sugimoto, and V. Hlaváč, “Multi-view facial landmark detector learned by the structured output SVM,” *Image and Vision Computing*, vol. 47, pp. 45–59, 2015.

- [56] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, <https://arxiv.org/abs/1207.0580>.
- [57] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the ICLR*, pp. 1–15, Banff, Canada, April 2014.
- [58] K. Ricanek Jr. and T. Tesafaye, "MORPH: a longitudinal image Age-progression, of normal adult," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGRO6)*, pp. 1–4, Southampton, UK, April 2006.
- [59] 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.