

# AGE ESTIMATION USING SPECIFIC DOMAIN TRANSFER LEARNING

Arwa Al-Shannaq<sup>1</sup> and Lamiaa Elrefaei<sup>2</sup>

(Received: 19-Oct.-2019, Revised: 11-Dec.-2019, Accepted: 29-Dec.-2019)

## ABSTRACT

Nowadays, the engagement of deep neural networks in computer vision increases the ability to achieve higher accuracy in many learning tasks, such as face recognition and detection. However, the automatic estimation of human age is still considered as the most challenging facial task that demands extra efforts to obtain an accepted accuracy for real application. In this paper, we attempt to obtain a satisfied model that overcomes the overfitting problem, by fine-tuning CNN model which was pre-trained on face recognition task to estimate the real age. To make the model more robust, we evaluated the model for real age estimation on two types of datasets: on the constrained FG\_NET dataset, we achieved 3.446 of MAE, while on the unconstrained UTKFace dataset, we achieved 4.867 of MAE. The experimental results of our approach outperform other state-of-the-art age estimation models on the benchmark datasets. We also fine-tuned the model for age group classification task on Adience dataset and our model achieved an accuracy of 61.4%.

## KEYWORDS

Age estimation, Transfer learning, Classification, Regression, VGGFace, Convolutional neural network.

## 1. INTRODUCTION

To characterize the human identity, different attributes can be derived from the facial image. Age is a crucial trait that can support other significant properties, such as fingerprint and iris, to get a more realistic system for the task of identification and verification of human identity [1]. Age estimation is related to the automatic process of predicting the real age as an exact age of a person or classifying the image into age groups represented by age range [2]. Therefore, estimating the real age of a person is much harder than just classifying a person to which age's category he belongs. Age estimation problem is considered as the most challenging facial task that is affected by various internal factors, such as gender, race and external factors, such as environment conditions and facial expressions. Recently, a growing interest is witnessed to automate age estimation systems and great efforts are made to enhance this challenging task.

Deep learning is one of the new technologies having been increasingly used in the field of computer vision. There is no doubt how deep learning technology outperforms the traditional algorithms of machine learning. Nonlinear features can be automatically extracted using Convolutional Neural Networks (CNNs) [3]. The power of CNNs is related to their capability of hierarchical learning for concepts across several layers. Despite the success of deep learning in many tasks, the accuracy of age estimation systems is still practically unacceptable.

Transfer learning [4] offers many benefits for deep learning-based models. It increases the training speed on new data. Furthermore, it requires less amount of data to train the model compared with training from scratch. Also, it improves the performance of the network. Generally, there are two types of transfer learning: general-domain pre-trained models where knowledge is transferred from a general task to a target unrelated task, whereas models learned on a specific domain are pre-trained on a related or similar task to the target task.

Face recognition and age estimation are different tasks. However, we can argue that they are correlated. While the recognition learning process was made for the VGGFace of the facial features and landmarks across a large number of images related to the same person in different conditions, this process can be

---

1. A. Al-Shannaq is with the Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. E-mail: aalshannaq@stu.kau.edu.sa  
 2. L. Elrefaei is with the Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia and with The Electrical Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt. E-mail: laelrefaei@kau.edu.sa, lamiaa.alrefaei@feng.bu.edu.eg

considered as an initial step for an age estimation system rather than using a random initialization of weights for the network. To gain the advantage of reusing specific domain pre-trained models, our model is considering the VGGFace [5] as the CNNs architecture. The aim of choosing this network is related to many reasons: Firstly, the state-of-art results that have been achieved by this model in face recognition task. Secondly, VGGFace has been pre-trained on a large database that contains 2.6 M images of 2.6 K people and this can overcome the overfitting problem when training the same model on small datasets on related task of face recognition. Thirdly, using a pre-trained model makes the training process faster while increasing the overall performance. Finally, age can be considered as a facial trait, thus the learning process will be easier than using general domain transfer learning.

In this paper, we prove that selecting a good online data augmentation helps improve the performance. Using a pre-trained model on a large dataset for face recognition task has a good impact on extracting related features to age and can prevent the overfitting problem. Moreover, combining the Global Average Pooling (GAP) with Fully Connected (FC) layers to generate the classifier increases its ability to perform the age estimation task. Finally, treating the age as a multi-classification problem is good when using a balanced distribution over the dataset's classes.

The remainder of this paper is divided into five sections as follows: Section 2 gives a brief overview of the related work on age estimation. The methodology with the baseline network is outlined in the third section. Section 4 displays the different experiments that have been conducted to achieve the best performance with related results. A discussion of the results is addressed in the fifth section. Our conclusions are drawn in the final section.

## 2. RELATED WORK

To estimate real age, two main steps should be followed. Firstly, the image representation should be extracted, then the feature vector is formalized. Secondly, an encoding algorithm is used to estimate the age [6]. Traditional methods carry out the process of age estimation by extracting the related features to age then formalizing the feature vector. Before representing the different facial details, the faces should be detected and aligned. Then, for the stage of extraction of the local and global features, a descriptor such as Active Appearance Model (AAM) [7], Biologically Inspired Feature (BIF) [8] or Local Binary Patterns (LBPs) [9] should be carefully chosen. The dimension of the resulted feature vector could be very high. Therefore, a reduction algorithm such as Principle Components Analysis (PCA) is used to reduce the dimension. The traditional techniques are classified into six groups according to [10]: anthropometric-based [11][12][13] that are based on the face geometry. To measure the geometric ratios from images correctly, faces should be in frontal view, since the computations of ratios from 2D images are sensitive. The second method is texture-based model [8]-[9]. Texture information can be directly calculated from images using pixel intensities. To extract these features, many effective descriptors such as BIF have been utilized in several works [14][15][16]. AAM [7] is another effective algorithm that combines shape and texture. The models are learned through the training process using a number of images. After that a parametric model for faces is generated by PCA. Another model is aging pattern subspace [17]-[18] which defines the aging pattern as a sequence of personal facial images belonging to the same person and finally the aging manifold [19]-[20], where the aging pattern can be commonly learned as a trend for several subjects at different ages. The features from different descriptors can be also fused to obtain a more robust system [21]-[22].

Since the development of deep learning technology, many researchers turned into replacing the traditional techniques with this technology. CNNs [3] have been widely used in computer vision field due to their effective learning ability for nonlinear features. Recent works on age estimation using deep learning [23][24][25] have only focused on developing their approaches using models that were pre-trained on ImageNet [26] dataset for object classification task. Other approaches [27][28][29] found that using a model pre-trained for a specific-domain related to age prediction task such as face recognition can achieve a better performance on age estimation. Yang et al. [30] trained their CNN model from scratch to derive different information from human faces, such as age, gender and race. Their results were lowest compared to previous works on age estimation which were based on traditional techniques for feature extraction. Instead of extracting the features from the top layer, Wang et al. [18] obtained the features from different layers. They enhanced the system by adopting the manifold algorithm with the basic model. Liu et al. [28] built a multi-path model and combined a VGGFace network with two

shallow VGG16 [31] networks. The feature vector of the three networks was normalized and fed to the age estimator. Hu et al. [32] learned the age from two pairs of images which belong to the same person. They found the age difference by using the Kullback-Leibler divergence as loss function. Rodríguez et al. [33] tried to place more attention on VGGFace [5] that is pre-trained on face recognition task. The model consists of two CNNs; the patch network which receives the low-resolution image and the attention CNN which is fed with the high-resolution image to place more attention on the important regions on the face.

Estimating real age can be implemented as a multi-classification task [34][35]. In this case, ages are represented as separated labels. Therefore, the classifier will predict the class of the age after the learning process using one of the available benchmarks annotated with age labels. On the other hand, we can use the regression algorithm [36]-[37] because of the nature of the human ages, as they are continuous values. A new algorithm which shows good results in age estimation is called ranking algorithm [38] [39] [40]. Instead of considering the age as a multi-class task, ranking algorithm converts the problem into different binary classification tasks and the resulted age ranks are the aggregation of the outputs from different classifiers. Rothe et al. [27] pre-trained the VGG16 on a large unconstrained dataset: IMDB\_WIKI [27] for real age estimation task. They considered the age as a multi-class problem with linear regression activation.

Antipov et al. [29] used VGG16 and pre-trained it on face recognition task. They used a soft classification to encode age. They found that pretraining on face recognition task is more suitable for age and gender classification than general task pre-training, while the strategy of multi-task pretraining is useful in case of training the model from scratch. To overcome the problem of sample imbalance, Li et al. [25] used AlexNet for feature extraction with a cumulative hidden layer. The main advantage of using a cumulative hidden layer is to learn the ages from faces with neighboring ages. Shang and Ai [41] separated the related features of aging into different groups by using clustering algorithm k-means++. They trained the network again for each group to estimate the final age for each subject. Later, Zhang et al. [42] used the DEX method [27] to estimate the real age. They improved the age estimation system by extracting the fine-grained features using the attention mechanism. A new loss function was proposed by [43] based on finding the age distribution based on the mean and variance of the ground-truth age.

Deep network can be utilized as a feature extractor. Duan et al. [15] extracted the features using CNNs. They combined classification and regression by firstly classifying the images according to age groups using Extreme Machine Learning (ELM) and then regressing the final value of the age using ELM regressor. Their model was a combination between classification and regression. Chang and Chen [38] used the scattering transform to extract the Gabor coefficients. They treated the age labels as a ranking algorithm, where the aggregation of the results from a series of binary classifiers performs the age ranks. Chen et al. [39] trained a set of CNNs on ordinal age labels. Different outputs were obtained and aggregated to predict the final age. Recently, Li et al. [44] extracted the features using CNNs and fed the BridgeNet which consists of local regressors and gated networks. The aim of using the gated networks was to weigh the regression results. Thus, the final age was calculated by taking the summation of all weights resulted from the local regressors.

The literature shows that using a pre-trained model on general task, such as on ImageNet, needs more efforts and deep networks to achieve reasonable results for age estimation task. A more powerful model is to pre-train the network on a specific task, such as face recognition or gender classification. Inspired from this consequence, we use a pre-trained model on face recognition that can increase the system ability to extract the related features of age.

### 3. METHODOLOGY

In deep learning, transfer learning technique can be defined as the process of reusing a model that has already been trained for a specific task to perform a similar or related task [4]. The aim of using a pre-trained model is to take the advantage from the features that have been extracted in the front layers instead of developing the model from scratch. Moreover, the computation time for training can be reduced while using a pre-trained model. Different policies can be followed while reusing pre-trained models:

1. Some late layers can be set as a trainable layer, so that their weights will be fine-tuned for the new task. This policy can be used when a dataset is available with plenty labels.
2. Freezing the convolutional base and adding classification layers in case that a small dataset is available which is similar to the source dataset that has been used in the pre-training stage.
3. Training the entire model or training from scratch which needs extra computational time and power with a very large dataset.

To build the classification layers, we can use one of the following approaches:

1. Adding an FC layer or a set of stacked FC layers followed by a Softmax activated layer for classification task or a linear activated layer for a regression task.
2. Adding GAP as proposed by Lin et al. [45] and connecting this layer directly to the output layer. The main concept of GAP is that it reduces the dimension of each tensor by taking the average of each feature map. For example, if we have a tensor with dimension  $(h \times w \times d)$ , GAP reduces the dimension from  $(h \times w \times d)$  to  $(1 \times 1 \times d)$  by taking the average of each feature map  $(h \times w)$ . Moreover, this layer has a similar affect as the FC layer except that it can avoid overfitting, since there are no parameters to optimize.

In this study, different approaches were investigated to reuse the base model to build the classifier by adapting deeper and wider schemes. The model was also examined in terms of the power of concatenating the GAP with FC layers to leverage performance. Classification and regression were both implemented to estimate real age, while for the age group task, classification was implemented.

### 3.1 Network Architecture

VGGFace-based VGG16 consists of eight convolutional layers and the classifier layers. On the classification block, there are two FC layers, where each layer has a 4096-dimensional output. An activation layer that has the rectification operator, such as Rectified Linear Activation Unit (ReLU), is added between these FC layers. After each block, a max pooling layer is added to down sample the feature map. The last layer represents the output layer with 2622 classes reflecting the number of subjects in the database. The activation in the output layer is chosen to be Softmax function for multi-class classification problem. The base model with the top connected layers pre-trained on face recognition task is shown in Table 1.

### 3.2 Adding Batch Normalization between FC Layers

A good technique that approved its efficiency in avoiding overfitting problem is using regularization. Inserting Batch Normalization (BN) [46] between the convolution layers will regularize and make the model more stable. BN layer takes the output of the preceding activation layer and normalizes it by subtracting the mini-batch mean and dividing by the mini-batch standard deviation. In other words, the normalizing transform aims to repair the means and the variances of layer inputs by adding two trainable parameters at each layer. Moreover, it reduces the network dependency on the initialization of each layer, which allows to use a higher learning rate.

### 3.3 Improving the Model with Online Data Augmentation

One of the most important techniques to enhance the performance and robustness of deep learning models is to train the neural network with a large amount of data. Unfortunately, most of the image-based applications have limited datasets or the conditions do not reflect the real-world scenarios under which images have been taken. Age estimation task is considered as a challenging computer vision task that is affected by many internal and external factors [28], [47]. There are no general patterns of aging for all humans. A more realistic age estimation system should be able to learn more irrelevant patterns of ages in different conditions.

Data augmentation [48] is a regularization technique used to feed the neural network with more synthetic images to reflect more realistic conditions and perspectives. By applying data augmentation, the network can avoid the overfitting problem that is caused by using small datasets. Different conditions can be applied to make additional modified images, such as translation, rotation, scaling, brightness, ...etc. In this work, we select online augmentation to be applied to the images during the training process as the

batches are passed. Thus, the training process will be quicker and there is no need to load the original data with the augmented data to the memory as in offline augmentation [48]. We diagnosed different transformation functions of online data augmentation to select the most appropriate one that enhances the performance of the model.

Table 1. Main architecture of VGGFace network.

	Layer Name	# of Filters	Feature Map	Stride
	input_1 (InputLayer)		(224, 224, 3)	
<b>Block 1</b>	conv1_1 (Conv2D)	64	(224, 224, 64)	3x3
	conv1_2 (Conv2D)	64	(224, 224, 64)	3x3
	pool1 (MaxPooling2D)		(112, 112, 64)	2x2
<b>Block 2</b>	conv2_1 (Conv2D)	128	(112, 112, 128)	3x3
	conv2_2 (Conv2D)	128	(112, 112, 128)	3x3
	pool2 (MaxPooling2D)		(56, 56, 128)	2x2
<b>Block 3</b>	conv3_1 (Conv2D)	256	(56, 56, 256)	3x3
	conv3_2 (Conv2D)	256	(56, 56, 256)	3x3
	conv3_3 (Conv2D)	256	(56, 56, 256)	3x3
	pool3 (MaxPooling2D)		(28, 28, 256)	2x2
<b>Block 4</b>	conv4_1 (Conv2D)	512	(28, 28, 512)	3x3
	conv4_2 (Conv2D)	512	(28, 28, 512)	3x3
	conv4_3 (Conv2D)	512	(28, 28, 512)	3x3
	pool4 (MaxPooling2D)		(14, 14, 512)	2x2
<b>Block 5</b>	conv5_1 (Conv2D)	512	(14, 14, 512)	3x3
	conv5_2 (Conv2D)	512	(14, 14, 512)	3x3
	conv5_3 (Conv2D)	512	(14, 14, 512)	3x3
	pool5 (MaxPooling2D)		(7, 7, 512)	2x2
<b>Classification block</b>	Flatten layer			
	Fully Connected (FC6)		Input size= 4096	
	RELU Activation			
	Fully Connected (FC7)		Input size= 4096	
	RELU Activation			
	Fully Connected (FC8)		Input size= 2622	
	Softmax Activation layer			

### 3.4 Fine-tuning VGGFace Model for Age Estimation

A single network of VGGFace is used and fine-tuned for real-age estimation. We propose two approaches to reuse the basic model:

**Approach-1:** In this approach, we keep the base convolutional layers with the top classification layers and remove the last Softmax activation layer. Then, the resulted feature map from FC8 layer is connected to extra FC layers. We freeze all the layers except the new additional one. The model is examined when adding different numbers of FC layers that have different numbers of neurons. The last layer is the output layer which is connected to the model as a dense layer. This approach is shown in Figure 1.

**Approach-2:** In this approach, we want to combine the GAP layer and the FC layers. We keep the base convolution layers and remove the top classification layers, then connect the last max pooling layer with GAP layer. The output of GAP is then fed to the FC layers. The model is examined when adding different numbers of FC layers that have different numbers of neurons. The last layer is the output layer which is connected to the model as a dense layer. Figure 2 shows the proposed structure of approach-2 to fine-tune VGGFace network on age estimation.

### 3.5 Classification vs. Regression

Age can be estimated using one of the age encoding algorithms. If the age is considered as a multi-classification problem, then each age will represent a single class. However, we can find a correlation

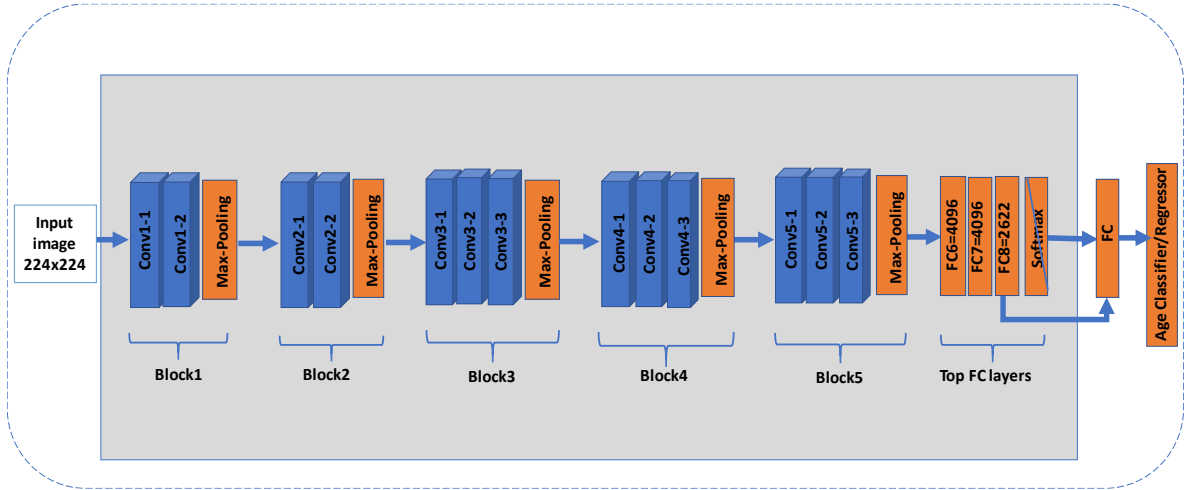


Figure 1. Approach-1 for fine-tuning VGGFace for age estimation.

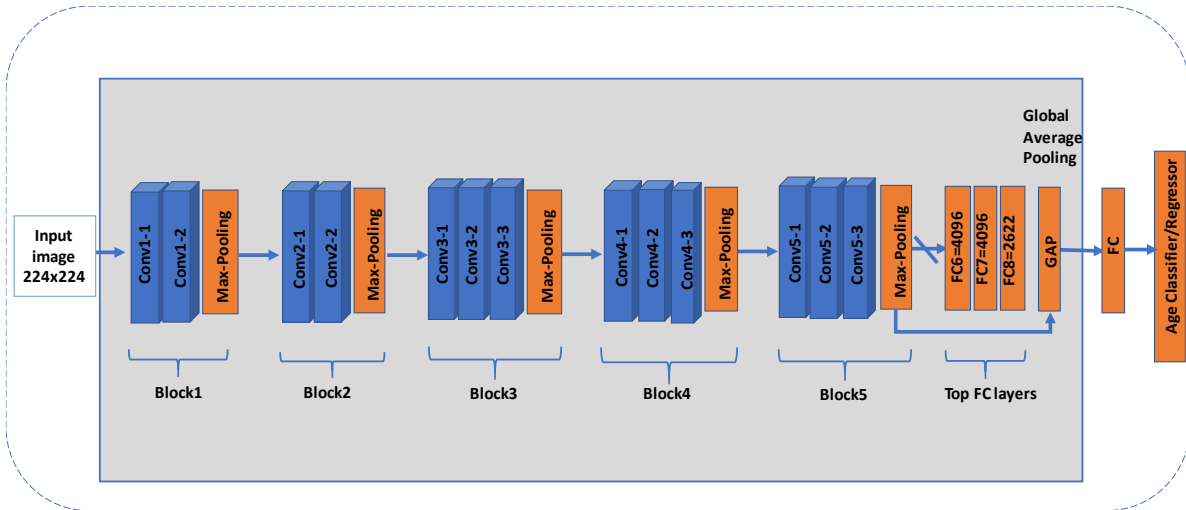


Figure 2. Approach-2 for fine-tuning VGGFace for age estimation.

and continuous behavior across different ages. In this case, ages can be considered as a regression problem. In this work, we investigated the classification and regression algorithms to encode age. By considering age as a multi-class problem, we added an age classifier on the top of the network as an output. We used the one-hot encoding method to represent the age labels. Thus, each sample will have a probability value of 1.0 for the correct class and 0.0 for other class values. To predict this probability, Softmax was applied as the activation function. Softmax function [49] can be defined as in Equation 1:

$$P(y_i | x_i, W) = \frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \quad (1)$$

where  $P$  is the probability value that is assigned to the correct label  $y_i$  given the image  $x_i$  and parameterized by  $W$ . Furthermore, to monitor the training process, categorical-cross entropy was selected as the loss function. The number of outputs in this case is equal to the number of classes on the dataset. Regression method treats the age as continuous values. In our proposed model, we added an age regressor with one output and selected the Mean Absolute Error (MAE) as a loss function to calculate the cost of training process. To estimate ages, we selected the linear activation function. A linear regression line can be represented using Equation 2:

$$Y = a + bX \quad (2)$$

where  $X$  is the explanatory variable and  $Y$  is the dependent variable. The slope of the line is  $b$  and  $a$  is the intercept (the value of  $y$  when  $x = 0$ ).

## 4. EXPERIMENTS AND RESULTS

We practically examined the efficiency of our proposed approaches to fine-tune a pre-trained network

of VGGFace for real-age estimation. We conducted many experiments to achieve the minimum value of error. The process of learning was further enhanced using online data augmentation, such as rotation, shearing and flipping. After trying many functions for image transformation, we selected the proper augmentation functions and considered them for all experiments. Thus, each image on the training set will be flipped horizontally, sheared randomly to 0.5 and rotated 45 degrees.

#### 4.1 Evaluation Metrics

The most common metric for real-age estimation is the MAE. This metric can be more representative than classification accuracy to evaluate the age estimation system, since it shows the difference between the estimated age and the real age. This metric can be mathematically defined as in Equation 3:

$$MAE = \frac{1}{N} \sum_{i=1}^n |\hat{x}_i - x_i| \quad (3)$$

The estimated age is represented by  $\hat{x}_i$ , where  $x_i$  represents the real age and  $N$  is the number of testing samples. The evaluation between the different models is based on obtaining a minimum value of  $MAE$ , which implies a good performance [2].

For the task of age group classification, accuracy is used as the evaluation metric. We calculated accuracy by using Equation 4:

$$CS(x) = \frac{n_x}{N_x} * 100\% \quad (4)$$

where  $n_x$  is the number of correctly classified images to a specific group  $x$  and  $N_x$  represents the number of testing samples of group  $x$  [17].

#### 4.2 Benchmark Datasets

**FG\_NET Dataset** [50] has 1002 images which are belonging to 82 subjects. There is a grayscale resolution for all images in addition to the colored version. The dataset covers a range of ages from 0 to 69 years, but most of the available images are in the range less than 40 years. On average, there are 12 images per person. The human gender and race annotations are also provided in this dataset. FG\_NET is considered as a constrained dataset, where all the images have frontal head poses captured on restricted conditions. Moreover, for the task of face modelling, the database provides 68 landmark points. Because of the problem of highly biased classes in this dataset, we considered a different protocol. Different recent age estimation models used the Leave One Person Out (LOPO), which depends on testing the model using the different images that belong to one person each time and taking the average obtained from all 82 subjects. Thus, to make the testing process more realistic, we considered another splitting protocol. The dataset is divided into 80 % as training set, 10% as validation set and 10% for testing. Figure 3 shows some samples of the dataset.



Figure 3. Samples of FG-Net dataset [50].

**UTKFace Dataset** [51] has over 20,000 face images labelled with age, gender and ethnicity. The images cover a large span of ages from 0 to 116 years. Huge variations in head poses, illumination and occlusion are contained in the images. It has images captured in unconstrained conditions with correct real-age annotations.

Figure 4 shows some images from the dataset. The dataset has been used to evaluate the model on an unconstrained dataset for real-age estimation using validation splitting of 80% for training, 10% for validation and 10% for testing.



Figure 4. Samples from UTKFace dataset [51].

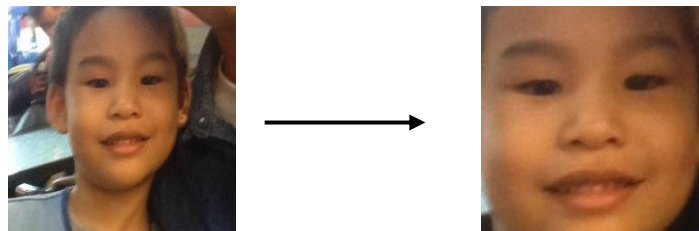
**Adience Dataset** [22][52]. It was collected from Fliker.com albums. It is used to make age group and gender classification. Some images from the dataset are shown in Figure 5. Adience dataset consists of 26,580 images that belong to 2,284 subjects. The set contains 8 age groups from 0 to 60 years and older. Big challenges are embodied in this database, while the images have low resolution with extreme blurring, occlusion, different head poses and expressions. Five models are trained for each fold. As the Adience dataset is already divided into five folders, four folders have been used for training and validation while folder0 has been used as a hold-out testing set. Then, accuracy is tested by the average or mean of all accuracies obtained from the five folds. The faces have been detected and the images were cropped to remove the annoying background.



Figure 5. Samples from Adience dataset [22], [54].

### 4.3 Data Pre-processing

For the Adience dataset, we detected the faces from the images and cropped them to remove the noisy background, as shown in Figure 6.



Original image

After face detection and cropping

Figure 6. Pre-processing step for images in Adience dataset.

For the UTKFace dataset, we used the aligned and cropped version of the dataset offered from the dataset's website. For the FG\_NET dataset, we used the images without making any processing. For all datasets, we rescaled the images to 224x224 resolution to be compatible with VGGFace input size.

### 4.4 Hardware and Software Tools

The system and its stages were executed using a High-Performance Computer (HPC) with NVidia Tesla GPU. Our model is implemented in Python using Keras with TensorFlow backend.

### 4.5 Experimental Results

This subsection summarizes the results that were obtained from the experiments for both encoding algorithms: classification and regression using the earlier proposed approaches. **Section 4.5.1** shows the results on the constrained FG\_NET dataset. In **section 4.5.2**, we show the results on the unconstrained UTKFace dataset using the proposed approaches. One of the main concerns that has a large effect on performance while reusing the pre-trained model is to optimize the appropriate topology of the classifier with strong generalization of the task. In our experiments, we studied different schemes to connect the pre-trained model with the output layers with changeable number of hidden layers (depth) and number



of neurons (width) to construct the most proper structure that can design the age estimation task. We analyzed the proposed approaches using the following different schemes:

- **Scheme-1:** We fine-tuned the base model without adding any FC layers. For Approach-1, we just connected FC8 to the output layer. For Approach-2, we just connected the GAP to the output layer.
- **Scheme-2,3,4:** Varying width, single layer with varying feature size. We started to expand the model by adding one layer with different neuron sizes, such as 1024, 5000 and 6000.
- **Scheme-5,6,7,8:** Varying depth, instead of using one layer, adding a set of two layers. Within this scheme, we examined many combinations. We combined BN with FC layers. After each BN layer, we added ReLU activation layer connected to the FC layer with the following order: FC1+BN+ ReLU +FC2+ BN+ ReLU. The reverse order was also tested: BN+ ReLU +FC1+ BN+ ReLU +FC2 to examine which order is more effective. Another scheme with FC layers and ReLU activation has also be considered, where each FC layer is followed with ReLU layer.

After conducting a number of experiments with different values of the training's parameters, we adjusted the values of these parameters for both datasets as follows: epochs=64, batch size=16, learning rate= 0.0001, split ratio=( 80 training,10 validation,10 testing) with Stochastic Gradient Descent (SGD) as an optimizer. To boost performance, we used online data augmentation.

#### 4.5.1 Results on the Constrained FG\_NET Dataset

In this section, we show the results of evaluating the proposed approaches on the constrained FG\_NET dataset. We used the images from the dataset without making any pre-processing. For the case of classification, the number of nodes for the output layer was set to 70, which is related to the number of classes in the dataset. For regression, the number of nodes was set to 1.

##### Approach-1: Fine-tuning the base convolutional layers with including top

The model was constructed by freezing the entire model and removing the Softmax activation layer. Then, we connected the feature vector from FC8 with different sizes of extra dense layers. Table 2 shows the results of classification and regression algorithms while connecting different schemes to the base model. It is clear from the results that, using Approach-1 with regression is better than using it with classification and resulted with a more robust model to age estimation. The least MAE was achieved when using the (BN+ ReLU+ Two FC (4096)) with adding BN and activation before each FC layer.

##### Approach-2: Fine-tuning the base convolutional layers without including top

The fine-tuning process was handled by freezing the base convolutional layers and removing the classification layers. We used GAP to connect the base convolutional layers with the new FC layers.

Table 2. Age estimation results on FG\_NET using approach-1.

Scheme Name	Additional FC layer	MAE of Classification	MAE of Regression
No FC	Without FC layer	4.222	5.455
FC1(1024)	FC1(input_size= 1024)	<b>3.839</b>	3.855
FC1(5000)	FC1(input_size= 5000)	4.041	3.894
FC1(6000)	FC1(input_size= 6000)	4.260	4.086
BN+ ReLU+Two FC (4096)	BN+ ReLU + FC1(input_size= 4096) + BN+ ReLU + FC2(input_size= 4096)	4.134	<b>3.446</b>
Two FC (4096) + BN+ ReLU	FC1(input_size= 4096) +BN+ ReLU + FC2(input_size= 4096) +BN+ ReLU	4.039	4.347
Two FC (4096) + ReLU	FC1(input_size= 4096) + ReLU + FC2(input_size= 4096) + ReLU	4.267	3.855
Two FC (4096)	FC1(input_size= 4096) + FC2(input_size= 4096)	4.092	3.976

Table 3 shows the results of Approach-2 for real-age estimation task on FG\_NET dataset using classification and regression to encode age. The lowest value of MAE was achieved when using (Two FC (4096) + ReLU) with regression algorithm.

Figure 7 shows the results of classification and regression encoding for different proposed schemes. We observed that, when using Approach-1, the regression algorithm on average obtained lower MAE than classification algorithm. We can see the good impact of using Approach-2 with the structure of stacked FC with activation layers (Two FC (4096) + ReLU) and without activation layer (Two FC (4096)). Reusing the model with (No FC) scheme was the worst case for both approaches. When using the scheme (BN+ ReLU+ Two FC (4096)), Approach-1 obtained the least MAE. On scheme (Two FC (4096) + ReLU), the second Approach got the least MAE. Better results of regression over classification when fine-tuning the model on FG\_NET are related to the biased classes on the dataset for the younger age and there is inadequate data to represent each class. Thus, learning the age as continuous values rather than discrete age labels is more efficient.

Table 3. Age estimation results on FG\_NET using approach-2.

Scheme Name	MAE of Classification	MAE of Regression
No FC	5.852	6.542
FC1(1024)	4.050	4.179
FC1(5000)	4.198	4.441
FC1(6000)	<b>3.807</b>	3.845
BN+ ReLU+ Two FC (4096)	5.300	6.184
Two FC (4096) + BN+ ReLU	5.934	5.048
Two FC (4096) + ReLU	3.837	<b>3.545</b>
Two FC (4096)	3.986	3.618

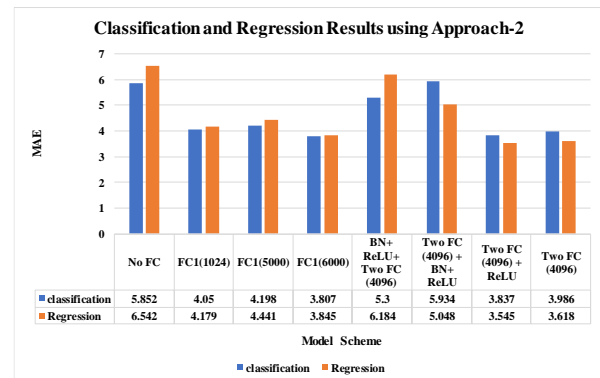
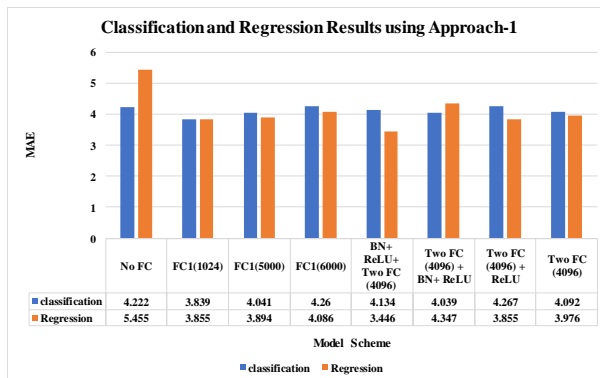


Figure 7. Age estimation results on FG\_NET using the proposed approaches.

#### 4.5.2 Results on the Unconstrained UTKFace Dataset

To make the model closer to real applications, we should evaluate it under rougher conditions. UTKFace is considered as an unconstrained dataset, where its images contain a diversity of head poses, illumination and occlusion. In this subsection, we show the results of evaluating the proposed approaches on the unconstrained UTKFace dataset. We used the aligned and cropped version of this dataset. For the case of classification, the number of nodes for the output layer was set to 101, which is related to the number of classes in the dataset. For regression, the number of nodes was set to 1.

Figure 8 shows the age estimation results when using Approach-1 and Approach-2 on UTKFace dataset. The results show the effectiveness of using classification algorithm with the proposed approaches to encode age. The lowest value of MAE was achieved by the scheme (Two FC (4096) + ReLU). It can be observed from the preceding experiments, how the performance of the model using Approach-2 outperforms that of Approach-1 for both encoding algorithms. Both approaches show an instable performance when using scheme (BN+ ReLU+ Two FC (4096)) and scheme (Two FC (4096) + BN+

ReLU) with regression algorithm, which is related to use a small batch size with large dataset, thus the BN may have an adverse effect. The least MAE value was achieved when using the FC layers with ReLU activation for both approaches. It is noticed that approach-1 shows more stable performance than approach-2 across all schemes.

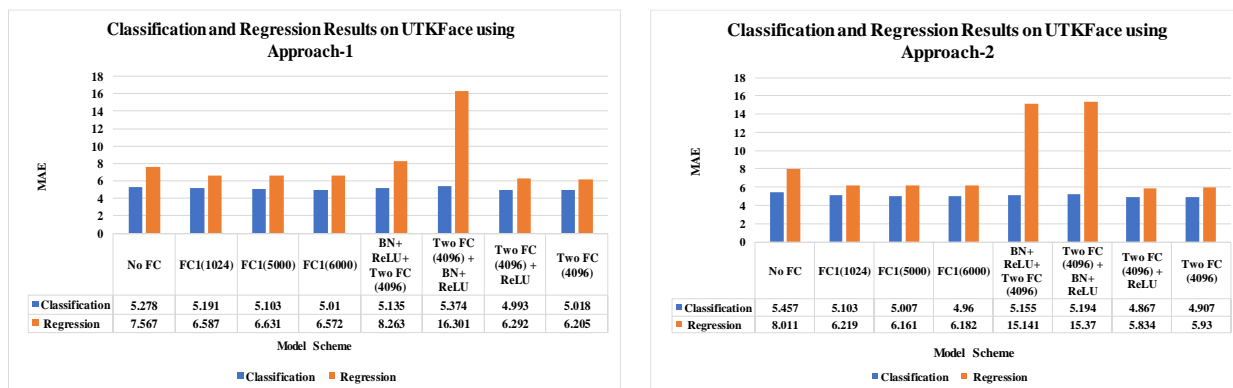


Figure 8. Age estimation results on UTKFace dataset using proposed approaches.

### 4.5.3 Results on the Unconstrained Adience Dataset

In this section, we show the results of reusing the base model with the second approach on Adience dataset for the age group classification task. In this experiment, the images were pre-processed by detecting and cropping the faces to remove the background. We fine-tuned the VGGFace using Approach-2 that showed a better performance than Approach-1, by connecting the GAP layer with different schemes to find the age group classification task. The output layer was set to 8 neurons related to the number of classes with Softmax as activation function. In this model, we considered accuracy as the evaluation metric. As shown in Figure 9, the classification for 8 classes of age groups needs a less complex, less deep network than the task of real-age estimation. The best accuracy was achieved by the base model without adding any extra layers, by directly connecting the GAP layer with the output layer.

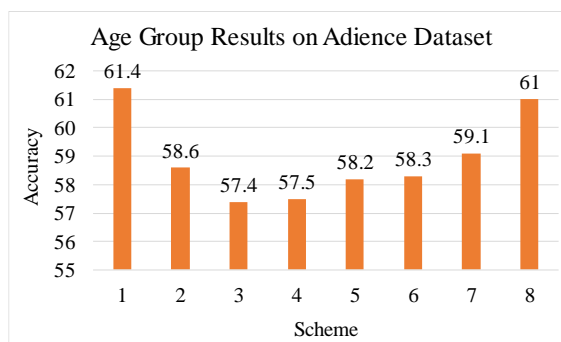


Figure 9. Age group classification results on Adience dataset using approach-2.

## 5. DISCUSSION

In this section, the proposed model is investigated according to some different criteria such as the effectiveness of specific domain transfer learning, the robustness and the complexity of the model. Moreover, some failure and success cases with more analysis are presented.

**The effectiveness of using specific domain transfer learning for age estimation system:** The typical structure of VGGFace is based on the pre-trained version of VGG16 network on VGGFace dataset for face recognition task. To show the efficiency of our proposed method, general domain VGG16 model pre-trained on ImageNet was tested. We tested VGG16 with approach-2 while using scheme (no FC). From the results shown in Table 4, we can see how approach-2 has profited from specific domain transfer learning compared with general domain for age estimation task.

We compare the pre-trained model of VGGface and VGG16 on how age estimation can benefit in extracting more features from face. The feature maps are visualized from the first block. Figure 10

shows the ability of the model VGGFace pre-trained on faces to consider and extract more features related to face. On the other side, we can see how VGG16 pre-trained on object task classification on the ImageNet dataset needs building a more complex network to be capable to extract more features related to face. From this point, we can conclude that using models that pre-trained on a task related to age such as face recognition is more effective than using models pre-trained on a general task.

Table 4. Results on FG\_NET using specific and general domain pre-trained models.

Pre-training Strategy	Model	MAE
Proposed model (Specific domain)	VGGFace-Approach-2	5.852
Pre-trained model (General domain)	VGG16-Approach-2	6.479

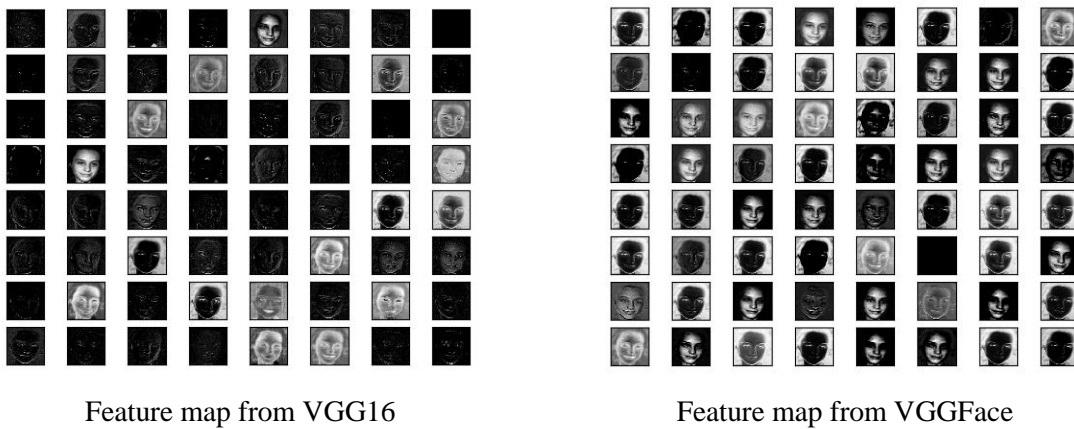
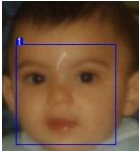



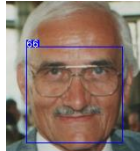



Figure 10. Feature's map from the second convolutional layer from VGG16 and VGGFace.

**The complexity of the model:** adding two FC layers means extra computational efforts and large number of parameters. Nowadays, the common direction tends to design less complex models that can be adapted to real application. For both datasets used, the system achieved reasonable results while using scheme (one FC with 6000 inputs) when encoding age using classification algorithm. In this case, the resulting models have lower sizes compared with models constructing using two FC layers, thus these models can be used in low memory devices that have a limited capacity, such as mobile applications.

**The robustness of the model:** using data augmentation in deep neural networks increases the ability of the network to learn from the same images with different conditions that were created using different transformation functions. Moreover, testing the model using only constrained datasets with clear data does not add any extra benefit to the system when using it in real application. Thus, using unconstrained datasets such as UTKFace and Adience that contain a large diversity of illusion, facial expressions and occlusion improves the system robustness to real conditions. Also, age estimation system needs a large number of images that cover a large range of ages to achieve more robust system.

**Success and failure cases of the model:** we tested our best model on real age estimation using different images from both datasets (FG\_NET, UTKFace). As shown in Figure 11, our model can successfully estimate the real age for both constrained and unconstrained images. In some cases, our system failed to find the real-age, where a person may look younger or older than his/her real age.

Successful Samples						
Real age	1	3	25	50	63	73
Predicted age	1	5	26	51	66	72

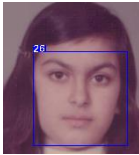
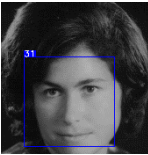


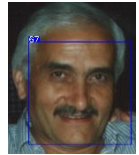

Failed Samples						
Real age	12	24	31	40	53	101
Predicted age	26	31	25	53	67	90

Figure 11. successful and failed samples tested by our best model.

## 6. COMPARISON WITH THE-STATE-OF-THE-ART

Corresponding to the earlier experiments, transfer learning from a pre-trained model on a specific domain, such as face recognition, can improve age estimation performance and overcome the overfitting problem. In addition, correct selection of data augmentation can enhance performance. In this section, a comparison of the results for the proposed model with the-state-of-the-art is described. Table 5 summarizes the results of age estimation on the constrained FG\_NET dataset with the-state-of-art. Some models such as [42] needed deeper network to reach reasonable results. Antipov et al. [29] benefited from pre-training on face recognition to improve the performance of real-age estimation. Chen et al. [39] pre-trained the ranking CNNs to extract the facial features on a small unconstrained dataset. Chang and Chen [38] used the hand-crafted features on their model to extract age features from facial images. A classification model for age estimation was proposed by Rothe et al. [27]. They used a VGG16 and pre-trained it on unconstrained IMDB\_WIKI dataset for age estimation task. IMDB\_WIKI dataset that contains about 500k images, which is a small number compared with the Face dataset that contains 2.6M images to pre-train the VGGFace.

Table 5. Results for age estimation on FG-NET dataset in various methods.

Ref.	Model	Base Model	Pre-training Strategy	MAE
Our proposed model	Regression_Approach-1 with Scheme:(BN+ ReLU + Two FC(4096))	VGGFace	Specific	<b>3.446</b>
	Regression_Approach-2 with Scheme:(Two FC(4096) +ReLU)	VGGFace	Specific	<b>3.545</b>
Zhang et al. (2018) [42]	Fine-Grained with attention mechanism	ResNets-34	Specific	2.39*
Li et al. (2019) [44]	BridgeNet (Local regressor + Gated network)	VGG-16	General	2.56*
Antipov et al. (2017) [29]	Soft classification + using LDAE	VGG-16	Specific	2.84*
Shang and Ai (2017) [41]	Feature Clustering using k-means++	GoogleNet	General	3.85*
Chen et al. (2018) [39]	Ranking-CNN	Shallow CNN	Training From scratch	4.13
Rothe et al. (2016) [27]	Classification + using DEX method	VGG16	Specific	4.63
Chang and Chen (2011) [38]	Ranking using Scattering Transform	Hand-crafted-based	-	4.48*

\*These models used a different testing protocol; LeaveOnePersonOut (LOPO).

The results show the effectiveness of using specific domain pre-training when extracting features related to age. Moreover, the combination of GAP and FC layers in Approach-2 with scheme (Two FC (4096) +ReLU) has regularized the model and prevented the overfitting problem. Using the scheme (BN+ ReLU + Two FC(4096)) has also a good influence on model stability.

Table 6 summarizes the results of age estimation on the unconstrained UTKFace dataset with other models. Niu et al. [53] constructed their model of multi-output CNN by considering less number of

layers than VGGFace. Cao et al. [54] used the VGG16 as the base network that was pre-trained on a general domain. It is clearly observed from the results that using VGGFace based on VGG16 for age estimation model is more efficient than using other models. For our proposed model, the least error was obtained when using classification as an age encoding algorithm while using Approach-2 for fine-tuning the base model with (Two FC (4096) +ReLU). It is noticed that our proposed model has obtained lower MAE when using classification than regression. This can be obviously related to that UTKFace is a large dataset with balanced distribution according to age. Thus, treating age as a multi-classification task will obtain good results if each class in the dataset has an adequate number of images with balanced distribution over a wide range of ages.

Table 6. Results for age estimation on UTKFace in various methods.

Ref.	Model	Base Model	Pre-training Strategy	MAE
Our proposed model	Classification_Approach-2 with Scheme: (Two FC (4096) +ReLU)	VGGFace	Specific	<b>4.867</b>
	Classification_Approach-1 with Scheme: (Two FC (4096) +ReLU)	VGGFace	Specific	<b>4.993</b>
Cao et al. (2019) [54]	CORAL-CNN: Ordinal Regression	VGG-16	General	5.83
Niu et al. (2016) [53]	Multiple Output CNN + Ordinal Regression	CNN	General	6.39

Table 7. Results of age group on Adience dataset in various methods.

Ref.	Model	Base Model	Pre-training Strategy	Accuracy
Our proposed model	Classification_Approach-2 with Scheme: No FC	VGGFace	Specific	<b>61.4</b>
Zhang et al. (2018) [42]	Fine-Grained + Attention	ResNets-152	Specific	67.83
Rothe et al. (2016) [27]	DEX method	VGG16	Specific	64.0
Rodríguez et al. (2017) [33]	Attention	VGGFace	Specific	61.8
Rodríguez et al. (2018) [55]	Attention	WRN	General	59.7
Chen et al. (2018) [39]	Ranking-CNN	Shallow CNN	Training From scratch	53.7
Levi and Hassner (2015) [52]	CNNs	Shallow CNN	Training From scratch	50.7

For age group classification task, as shown in Table 7, we achieved a good accuracy in spite of the simple fine-tuning scheme and the approach that were used in this experiment. Although Zhang et al. [42] achieved a higher accuracy than other methods, they used a very deep residual network with 152 layers compared with less deep network such as VGG with 16 layers which was the base model for [27] and was pre-trained on a large dataset. Rodríguez et al. [55] used a Wide Residual Network (WRN) pre-trained on a general domain task. For their previous work [33], despite adding the attention mechanism to improve the performance of VGGFace, they achieved a minor enhancement in accuracy compared with our model using the GAP layer that can reduce the overfitting problem. Other models just used shallow CNNs to extract age features.

## 7. CONCLUSIONS

This paper proposes an age estimation model based on VGGFace, which was pre-trained on a specific domain. Age is an attribute that is derived from face, thus reusing a model pre-trained on a related task

to age can have the capability to extract discriminative features related to age effectively and avoid the overfitting problem when using limited data. In this work, we utilized different approaches for fine-tuning the basic VGGFace model. The first approach kept the top classification layers with the base convolutional layers and connected the model with different strategies of adding extra FC layers. The second approach joined the base convolutional layers with the GAP connected to different strategies of adding extra FC layers.

The proposed approaches were tested under different schemes, by varying feature size and model depth. We also investigated the effectiveness of two algorithms for encoding age: classification and regression. A minimum error was obtained when using a balanced distributed dataset like UTKFace, where each class is represented by enough data at a specific age. On the other hand, in case of unbalanced datasets like FG\_NET, regression performance was better than classification performance. Furthermore, selecting appropriate data augmentation can improve performance, such as rotation, shearing and flipping. We evaluated our model using two kinds of datasets: constrained FG\_NET and unconstrained UTKFace and Adience. Our model achieved state-of-the-art results on FG\_NET when using regression to encode age, while the lowest MAE was obtained when using classification for UTKFace dataset. According to the age group classification task, the model was fine-tuned using the second approach and good results on Adience dataset were achieved despite the simple approach of fine-tuning that was implemented. The small degree of complexity to classify 8 age groups compared to large age range could explain the good performance that was obtained.

For future work, we will try to use a hybrid system that combines classification and regression in one model. The idea is to firstly classify the images into age groups, then for each group, a regression CNN will be trained to estimate age.

## ACKNOWLEDGEMENTS

Computations for the work described in this paper were supported by King Abdulaziz University's High-Performance Computing Centre (Aziz Supercomputer) (<http://hpc.kau.edu.sa>).

## REFERENCES

- [1] A. K. Jain, S. C. Dass and K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems," Proceedings of International Conference on Biometric Authentication, Berlin, Heidelberg, vol. 3072, pp. 731–738, 2004.
- [2] Y. Fu, G. Guo and T. S. Huang, "Age Synthesis and Estimation via Faces: A Survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 11, pp. 1955–1976, 2010.
- [3] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
- [4] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, 2010.
- [5] O. M. Parkhi, A. Vedaldi and A. Zisserman, "Deep Face Recognition," British Machine Vision Conference, 2015.
- [6] T. Zheng, W. Deng and J. Hu, "Age Estimation Guided Convolutional Neural Network for Age-Invariant Face Recognition," Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 503–511, 2017.
- [7] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active Appearance Models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 681–685, 2001.
- [8] G. Guo, T. S. Huang, Y. Fu and T. S. Huang, "Human Age Estimation Using Bio-inspired Features," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, USA, pp. 112–119, 2009.
- [9] T. Ahonen, A. Hadid and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, 2006.
- [10] A. S. Al-Shannaq and L. A. Elrefaei, "Comprehensive Analysis of the Literature for Age Estimation From Facial Images," IEEE Access, vol. 7, pp. 93229–93249, 2019.
- [11] Y. H. Kwon and N. da V. Lobo, "Age Classification from Facial Images," Comput. Vis. Image Underst., vol. 74, no. 1, pp. 1–21, Apr. 1999.

- [12] A. Gunay and V. V. Nabiyev, "Automatic Detection of Anthropometric Features from Facial Images," Proc. of the 15<sup>th</sup> IEEE Signal Processing and Communications Applications, pp. 1–4, 2007.
- [13] M. M. Dehshibi and A. Bastanfard, "A New Algorithm for Age Recognition from Facial Images," Signal Process., vol. 90, no. 8, pp. 2431–2444, Aug. 2010.
- [14] G. Guo and G. Mu, "A Framework for Joint Estimation of Age, Gender and Ethnicity on a Large Database," Image Vis. Comput., vol. 32, no. 10, pp. 761–770, 2014.
- [15] G. Guo and G. Mu, "Simultaneous Dimensionality Reduction and Human Age Estimation *via* Kernel Partial Least Squares Regression," CVPR 2011, pp. 657–664, 2011.
- [16] H. Han, C. Otto and A. K. Jain, "Age Estimation from Face Images: Human vs. Machine Performance," Proc. of the 2013 Int. Conf. Biom. (ICB), pp. 1–8, 2013.
- [17] X. Geng, Z. Zhou, S. Member, K. Smith-miles and S. Member, "Automatic Age Estimation Based on Facial Aging Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 12, pp. 2234–2240, 2007.
- [18] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li and H. Dai, "Learning from Facial Aging Patterns for Automatic Age Estimation," Proceedings of the 14<sup>th</sup> ACM International Conference on Multimedia, pp. 307–316, 2006.
- [19] Y. Fu and T. S. Huang, "Human Age Estimation with Regression on Discriminative Aging Manifold," IEEE Trans. Multimed., vol. 10, no. 4, pp. 578–584, 2008.
- [20] 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.
- [21] 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 Recognit., vol. 44, no. 6, pp. 1262–1281, 2011.
- [22] E. Eiding, R. Enbar and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," IEEE Trans. Inf. FORENSICS Secur., vol. 9, no. 12, pp. 2170–2179, 2014.
- [23] Z. Hu, Y. Wen, J. Wang and M. Wang, "Facial Age Estimation with Age Difference," IEEE Trans. IMAGE Process., vol. 7149, no. c, pp. 1–11, 2016.
- [24] K. E. Zhang, C. E. Gao, L. Guo, M. Sun and S. Member, "Age Group and Gender Estimation in the Wild with Deep RoR Architecture," Proc. of Chinese Conference on Computer Vision (CCCV), vol. 5, no. Cccv, 2017.
- [25] K. Li, J. Xing, W. Hu and S. J. Maybank, "D2C : Deep Cumulatively and Comparatively Learning for Human Age Estimation," Pattern Recognit., vol. 66, no. July 2016, pp. 95–105, 2017.
- [26] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis. (IJCV), vol. 115, no. 3, pp. 211–252, 2015.
- [27] R. Rothe, R. Timofte and L. Van Gool, "Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks," Int. J. Comput. Vis., vol. 126, no. 2–4, pp. 144–157, 2016.
- [28] H. Liu, J. Lu, J. Fenga and J. Zhou, "Group-aware Deep Feature Learning for Facial Age Estimation," Pattern Recognit., vol. 66, pp. 82–94, 2016.
- [29] G. Antipov, M. Baccouche, S. Berrani and J. Dugelay, "Effective Training of Convolutional Neural Networks for Face-based Gender and Age Prediction," Pattern Recognit., vol. 72, pp. 15–26, 2017.
- [30] M. Yang, S. Zhu, F. Lv and K. Yu, "Correspondence Driven Adaptation for Human Profile Recognition," Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 505–512, 2011.
- [31] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Proc. of the International Conference on Learning Representations (ICLR), vol. abs/1409.1, 2015.
- [32] Z. Hu, Y. Wen, J. Wang and M. Wang, "Facial Age Estimation with Age Difference," IEEE Trans. Image Process., vol. 7149, no. c, pp. 1–11, 2016.
- [33] P. Rodríguez, G. Cucurull, J. M. Gonfaus, F. X. Roca and J. González, "Age and Gender Recognition in the Wild with Deep Attention," Pattern Recognit., vol. 72, pp. 563–571, Dec. 2017.



- [34] 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, Feb. 2004.
- [35] K. Ueki, T. Hayashida and T. Kobayashi, "Subspace-based Age-group Classification Using Facial Images Under Various Lighting Conditions," *Proc. of the 7<sup>th</sup> IEEE International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 1-6, 2006.
- [36] A. Lanitis, C. J. Taylor and T. F. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, 2002.
- [37] Y. Fu, Y. Xu and T. S. Huang, "Estimating Human Age by Manifold Analysis of Face Pictures and Regression on Aging Features," *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 1383–1386, 2007.
- [38] K. Chang and C. Chen, "A Learning Framework for Age Rank Estimation based on Face Images with Scattering Transform," *IEEE Trans. Image Process.*, vol. 7149, no. c, pp. 1–14, 2015.
- [39] S. Chen, C. Zhang and M. Dong, "Deep Age Estimation: From Classification to Ranking," *IEEE Trans. Multimed.*, vol. 20, no. 8, 2018.
- [40] K. Li, J. Xing, C. Su, W. Hu, Y. Zhang and S. Maybank, "Deep Cost-sensitive and Order-preserving Feature Learning for Cross-population Age Estimation," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT*, pp. 399–408, 2018.
- [41] C. Shang and H. Ai, "Cluster Convolutional Neural Networks for Facial Age Estimation," *Proceedings of International Conference on Image Processing (ICIP)*, pp. 1817–1821, 2017.
- [42] K. Zhang, N. Liu, X. Yuan, X. Guo, C. Gao and Z. Zhao, "Fine-grained Age Estimation in the Wild with Attention LSTM Networks," *CoRR*, vol. abs/1805.1, pp. 1–12, 2018.
- [43] H. Pan, H. Han, S. Shan and X. Chen, "Mean-variance Loss for Deep Age Estimation From a Face," Presented at the *IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT*, pp. 5285–5294, 2018.
- [44] W. Li, J. Lu, J. Feng, C. Xu, J. Zhou and Q. Tian, "BridgeNet: A Continuity-aware Probabilistic Network for Age Estimation," *ArXiv190403358 Cs*, Apr. 2019.
- [45] M. Lin, Q. Chen and S. Yan, "Network In Network," *ArXiv13124400 Cs*, Dec. 2013.
- [46] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *ArXiv150203167 Cs*, Feb. 2015.
- [47] M. Duan, K. Li and K. Li, "An Ensemble CNN2ELM for Age Estimation," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 3, pp. 758–772, 2018.
- [48] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019.
- [49] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York, Springer, 2006.
- [50] A. Lanitis and T. Cootes, "Fg-net Aging Data Base," *Cyprus Coll.*, 2002.
- [51] Zhang Zhifei, Song, Yang and H. Qi, "Age Progression/Regression by Conditional Adversarial Autoencoder," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [52] G. Levi and T. Hassner, "Age and Gender Classification using Convolutional Neural Networks," *CVPR*, vol. 24, no. 3, pp. 2622–2629, 2015.
- [53] Z. Niu, M. Zhou, L. Wang, X. Gao and G. Hua, "Ordinal Regression with Multiple Output CNN for Age Estimation," *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 4920–4928, 2016.
- [54] W. Cao, V. Mirjalili and S. Raschka, "Rank-consistent Ordinal Regression for Neural Networks," *ArXiv190107884 Cs Stat*, Jan. 2019.
- [55] P. Rodríguez, J. M. Gonfaus, G. Cucurull, F. X. Roca and J. Gonzàlez, "Attend and Rectify: a Gated Attention Mechanism for Fine-Grained Recovery," *ArXiv180707320 Cs*, Jul. 2018.

**ملخص البحث:**

في الوقت الحاضر، تزيد مشاركة الشبكات العصبية العميقة في رؤية الحاسوب من القدرة على تحقيق دقة أعلى في العديد من مهام التعلم مثل التعرف على الوجوه والكشف عنها. ومع ذلك، لا يزال التقدير التلقائي لعمر الإنسان مهمة الوجه الأكثر تحدياً والتي تتطلب بذل جهود إضافية للحصول على دقة مقبولة للتطبيق الحقيقي. في هذه الورقة، نحاول الحصول على نموذج مُرضٍ يتغلب على مشكلة التهيئة الزائدة (overfitting)، من خلال ضبط نموذج شبكات عصبية التفاضلية (CNNs) تم اختباره مسبقاً في مهمة التعرف على الوجوه لتقدير العمر الحقيقي. لجعل النموذج أكثر قوة، قمنا بتقييم النموذج الخاص بتقدير العمر الحقيقي على نوعين من مجموعات البيانات: علي مجموعة البيانات FG\_NET المقيدة حققتنا 3.446 MAE، وعلى مجموعة البيانات UTKFace غير المقيدة حققتنا MAE 4.867. وتتفوق النتائج التجريبية لنهجنا في نماذج تقدير العمر الحديثة على مجموعات البيانات المرجعية. وقمنا بضبط نموذج مهمة تصنيف مجموعات العمر على مجموعة البيانات Adience، ونموذجنا حقق نتيجة جيدة.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).