Image and Vision Computing 88 (2019) 41–51

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images☆

Arafat Abu Mallouh [a], Zakariya Qawaqneh [b], Buket D. Barkana [c,*]

[a] Computer Science Department, Manhattan College, Riverdale, NY 10471, USA
[b] Department of Computing Sciences, The College at Brockport State University of New York, Brockport, NY 14420, USA
[c] Electrical Engineering Department, University of Bridgeport, Bridgeport, CT 06604, USA

ARTICLE INFO

Article history:
Received 16 March 2017
Received in revised form 3 February 2019
Accepted 2 May 2019
Available online 9 May 2019

Keywords:
Age range classification
CNNs
Deep learning
Deep neural networks (DNNs)
Face recognition

ABSTRACT

Automatic age classification from real-world and wild face images is a challenging task and has an increasing importance due to its wide range of applications in current and future lifestyles. As a result of increasing age specific human-computer interactions, it is expected that computerized systems should be capable of estimating the age from face images and respond accordingly. Over the past decade, many research studies have been conducted on automatic age classification from face images. However, the performance of the developed age classification systems suffered due to the absence of large, comprehensive benchmarks. In this work, we propose and show that pre-trained CNNs which were trained on large benchmarks for different purposes can be retrained and fine-tuned for age range classification from unconstrained face images. Also, we propose to reduce the dimension of the output of the last convolutional layer in pre-trained CNNs to improve the performance of the designed CNNs architectures. The experimental results show significant improvements in exact and 1-off accuracies on the Adience benchmark.

1. Introduction

The human face contains various information such as age, gender, emotional state, pose, and ethnic background. Such information could be extracted and used in entertainment, cosmetology, biometrics, human-computer interaction (HCI), security control, and surveillance monitoring applications. Recent developments in computer technology have a direct impact on the growth of the image processing techniques while enriching the applications in computer vision and graphics fields further. One of the most popular research fields which have gained attention over the years is the automatic estimation of age information from facial images.

The estimation of age from 2D images gains importance in many present and future applications including education, advertisement, merchandise, controlled media access, electronic health applications, information forensics and security [1] and more. At present, age appropriate education, ads, and merchandises can be offered to users [2]. Media content can be made available based on the user's age [3]. Research in artificial intelligence is rapidly developing. In near future, a computerized system called robot doctors may determine the correct dose of medicine for a patient depending on his/her age. Smart robots may select the right age appropriate attitude and language while socializing with humans.

Age classification is defined as determining the exact age or the age range of a person using 2D facial images [4,5]. There are several challenges in age classification. One of the main challenges is that people do age at variable rates that are affected by factors such as, genetic factors, social conditions, and life style. Some people can look years younger than their chronological age while some can look years older. Another challenge is the dissimilarity between aging rates of men and women. Wearing makeup and accessories either to look younger or to hide aging marks is another challenge [6]. With these challenges are in mind, the extraction of distinctive features in age classification from 2D images is a non-trivial task [7].

Recently, deep convolutional neural networks (CNNs) have shown remarkable performance in various computer vision fields, which are object recognition, face detection, and human pose estimation. CNNs attempt to mimic the human brain's visual cortex. The architecture of a typical CNN is composed of multiple convolutional layers and each layer consists of neurons [8]. The configuration settings of CNNs are problem dependent. Generally, each layer works on the output of the previous layer and tries to extract more abstract and discriminative features from the input data. Since the number and the size of CNNs parameters are very large, pooling and stride techniques are used to reduce the time needed for training.

☆ This paper has been recommended for acceptance by S. Todorovic

* Corresponding author.
E-mail addresses: aabumallouh01@manhattan.edu (A. Abu Mallouh), zqawaqneh@brockport.edu (Z. Qawaqneh), bdbarkana@aol.com (B.D. Barkana).

https://doi.org/10.1016/j.imavis.2019.05.001
0262-8856/

Motivated by the success of CNNs architectures in different fields, we use CNNs as feature extractors for automatic age range classification. Existing benchmarks for age classifications are relatively small compared to the benchmarks used in face recognition. Training a deep neural architecture using a small benchmark is problematic since training a very deep CNN architecture on a relatively small benchmark is liable to a critical overfitting. To overcome this problem, we use deep CNN architectures that are trained for other classification tasks such as image classification, semantic segmentation, and face recognition on large benchmarks. Then, these architectures are adapted and fine-tuned to estimate the age of a subject from an unconstrained 2D image.

This paper makes the following contributions.

- It shows that previously trained CNNs for different task on large benchmarks can be employed for age range classification successfully.
- It shows that features extracted from pre-trained models for domain specific tasks can be successfully used to improve the age classification task. Age classification benchmarks are relatively small compared to the benchmarks of face recognition task. One of the most challenging problems in the machine learning is the overfitting problem that occurs when using small benchmarks. The over fitting problem can be overcome by employing a pre-trained CNN on a large benchmark from related domain specific tasks.
- Two CNN architectures and network configurations are presented for training and prediction in age range classification.
- Extensive performance evaluation with/without dimensionality reduction is reported by comparing six pre-trained CNN architectures (VGG-Face, GoogLeNet, ResNet-50, VGG-VD-16, VGG-VD-19, and FNC-8s) from related domain specific tasks and their combinations for age range classification.
- This work proves that not only the number of training images and the number subjects in a training benchmark affect the performance of age range classification, but also the pre-training task of the employed CNN determines the network's performance.

The paper is organized as follows. Section 2 provides an overview of the previous works. Section 3 presents the benchmark. Section 4 covers the CNN architectures and training for age range classification. Dimensionality reduction is presented. Section 5 reports the experimental results and discussions. Finally, the conclusion is stated in Section 6.

2. Previous works

Over the last ten years, many studies have been carried out for the age estimation task from real-world and wild facial images. In this section, existing benchmarks are presented, and a brief review of the most significant and milestone works is given with regard to feature extraction and classification methods.

Kwon and da Vitoria Lobo [9] carried out one of the earliest works in age estimation by using facial images. Cranio-facial changes in feature-position ratios and skin wrinkles were used as features for three age groups (baby, young adult, and senior adult). Facial features were detected and their ratios were computed. Skin wrinkle analysis was performed. This early work in 1994 has shown that computing ratios and detecting the presence of wrinkles can yield the age from facial images. The same year, Farkas [10] presented a mathematical model to estimate the growth of a person's head from infancy to adulthood. This model was used to estimate the age of a person from a facial image. The drawback of this model is that the performance of the age estimation degrades for adults by using models which are built by using 2D images. One of the earliest research works in age estimation is based on face anthropometry. Face anthropometry is a science that deals with measuring sizes and proportions on human face. In general, the estimation of age from facial images using anthropometry model is limited to young ages. The shape of the human head does not change significantly in the adult years. Moreover, the ratio of distances for face geometry is calculated using 2D images but 2D images are sensitive to head pose. As a result, frontal face images are the only images that can be used to measure geometry of the face. Therefore, anthropometry is not suitable for age estimation by using real world and wild facial images.

Other approaches have been proposed based on the facial features (appearance model or face descriptor). Local and global facial features were extracted [11,12]. Texture and shape features were calculated by using a semantic-level description of the face to describe facial features. They also built a classification system to estimate different age groups with five year intervals. Their system was tested on a Japanese database of 500 subjects aging from 15 to 64 years old. Moreover, gender estimation was done to improve the performance of the age estimation. Since women and men have different aging rates, the inclusion of gender estimation enhanced the age estimation.

Ramanthan and Chellappa [13] worked on age progression in young face images and computed eight ratios of distance measures for modeling age progression. They proposed a craniofacial growth model by illustrating how the age-based anthropometric constraints on facial proportions translate into linear and non-linear constraints on facial growth parameters and proposed methods to compute the optimal growth parameters. The purpose of their work was to predict one's appearance across the years and to perform face recognition. Anthropometrical changes of human face and its size, shape, and textural patterns may adequate to estimate an individual's age up to the adult years.

Lanitis et al. [14] studied the aging effects on face images and described how the effects of aging on facial appearance can be explained. They built a statistical-based face model. By the proposed shape intensity face model and automatic age simulation, statistically significant improvement in the performance of the age classification system was reported.

Geng et al. [15,16] modeled the aging pattern as the sequence of an individual's face images sorted in time order by constructing a representative subspace. The aging pattern subspace (AGES) model built an aging pattern for different age stages. In case, the images of some ages were not available, they were synthesized by using EM-like iterative learning algorithm. The AGES was evaluated on the FG-NET database with a mean absolute error of 6.77 years. They reported that the performance of the model was significantly better than the existing age estimation methods and was comparable to that of the human observers. One of the limitations of AGES approach is that it assumes the availability of images representing the different ages of an individual. If images for different ages are not available, AGES assumes that there is an age pattern similar to the input image. AGES approach utilizes the Active Appearance Model (AAM) to calculate the face representation to encode the wrinkles of the face. AAM only encodes the image intensities which cannot describe the local texture information. Local texture information is important to represent the wrinkles of elderly people.

Manifold analysis is used and proved to be promising in age estimation from face images by several studies [17–19]. An age estimation framework was proposed by Fu et al. [17] using manifold analysis and learning methods to find a sufficient low-dimensional embedding space. Manifold data points were modeled with a multiple linear regression function. Age manifold model is more flexible than AGES, since images could be built using different person's images for unavailable images of some ages. Age manifold built the common aging pattern using the manifold embedding technique to learn the low dimensional aging trend from a group of face images for each age. Scherbaum et al. [18] also proposed a statistical age estimation method using manifold learning over a 3D morphable model.

Gunay and Nabiyev [20] used effective texture descriptor for appearance feature extraction and utilized local binary patterns (LBP) in automatic age estimation system. Using the nearest neighbor classification, their system achieved 80% accuracy on the FERET database. By using

AdaBoost, they achieved 80-90% of accuracy on the FERET and PIE databases. Gao and Ai [21] used the Gabor feature with fuzzy-LDA for age estimation. Their work showed that Gabor feature is more effective than LBP. Yan et al. [22,23] employed spatial flexible patches (SFP) to be the feature descriptor in order to handle images with small undesirable defects such as occlusions and head pose. They achieved mean absolute error (MAE) accuracy of 4.94 years on the FG-NET database. Sparse feature design, graphical facial features topology, geometry, and configuration, were proposed and age estimated based on the multiresolution hierarchical face model by artificial neural network (ANN) [24]. This system achieved MAE of 5.974 years on the FG-NET database. In [25], Mu et al. proposed bio-inspired features (BIF) for age estimation. The bio-inspired features have the ability to handle small rotations and scale changes effectively. By using the BIF with an SVM classifier, their work achieved MAE of 4.77 years on the FG-NET database. In [26], two feature sets, BIF and the age manifold were used. By using an SVM classifier, the system achieved MAEs of 2.61 and 2.58 years for female and male on the YGA database, respectively.

Literature has developed over time to enhance the performance of age estimation from face images. Naturally, each method tried to overcome the limitations of the previous methods by widening the range of domains. All the previously mentioned methods showed conditional significant performance where the used databases were either small on size or constrained to specific kind of images such as frontal and aligned pose images. On the other hand, the proposed work is applied on a larger database with unconstraint face images.

In [27], age estimation on real-life faces acquired in unconstrained conditions was studied. The local binary patterns (LBP) and Gabor features were exploited as face representation. Adaboost was used to learn the discriminative LBP-Histogram bins for age estimation. They achieved 55.9% of accuracy on the (Group Photos benchmark) by an SVM classifier. Alnajar et al. [28] adopted a learning-based encoding method for age estimation under unconstrained imaging conditions. Multiple codebooks for individual face patches were extracted and learnt. The orientation histogram of local gradients as the feature vector for code learning was used. An unconstrained database (Group Photos benchmark) which contains 2744 images was used and they achieved an absolute improvement of 3.6% over the study in [27] on the same database.

Recently, new benchmarks have been designed for the task of age estimation from face images. These new benchmarks are more challenging than the previous benchmarks in terms of quantity and quality. The size of the new benchmarks is much larger and most importantly the quality of the included images is categorized as unconstrained. The unconstrained images reflect the real world wild environments and are collected from online image repositories. The Group Photos [3] and the Adience benchmarks [1] are examples of these new benchmarks. Currently, the Adience benchmark is considered to be the newest and the most challenging benchmark for age and gender estimation from face images.

Recently CNNs and DNNs have been started to be used for feature extraction and classification for the facial age estimation due to their success in several computer vision fields. Levi and Hassner [6] used CNNs in age estimation for the first time. A simple CNN architecture was used as a feature extractor and a classifier to avoid overfitting problem. Ranjan et al. [29] proposed a cascaded classification and regression system based on a coarse age classifier. They introduced an age regressor for each age group based on the features extracted from the coarse age classifier. Then they used an error correcting method for correcting the regression errors for subjects. Chen et al. [30] proposed a system that the features were extracted from a pre-trained CNN on face identification. The extracted features were fed to a small neural network to regress the age of the subject. Yang et al. [31] proposed a generic deep network model that extracted facial features by using a convolutional scattering network. The dimension of these features was reduced by PCA. They estimated the age using three fully connected layers that act as category-wise rankers. Yi et al. [32] extracted local aligned patches using several facial landmarks. For each face image, 23 patch pairs were extracted in total. Each patch was trained in separate CNN and their final fully connected layer outputs were fused to estimate the age of a person. Liu et al. [33] proposed the AgeNet model to estimate the age apparent for the ChaLearn 2015 Apparent Age database. Two different CNN models were trained and fused to estimate the apparent age. Qawaqneh et al. [34] proposed a new model to jointly fine-tune two DNNs based on a new cost function. The two DNNs were trained on different feature sets, which were extracted from the same input data.

After ChaLearn LAP2015 [35] apparent age estimation dataset has been launched, several new methods for apparent age estimation based on DNNs have been proposed. Rothe et al. [36] proposed the Deep EXpectation (DEX) model. This model relied on the usage of deeper neural networks, usage of larger and diverse datasets for training, efficient alignment of the object in the input image, and the utilization of pre-trained networks for comparable inputs. Basically, the proposed architecture used the deep VGG-16 architecture which was pre-trained for image classification. Ranjan et al. [37] proposed an approach for age estimation from unconstrained images based on deep convolutional neural networks (DCNN). In general, their work is divided into four main steps: 1) face detection, 2) face alignment, 3) DCNN-based feature extraction and 4) neural network regression for age estimation. In their work they extracted the age-related features using a CNN trained for face identification over the CASIA-WebFace dataset.

The work in this paper is different from the previous works in several aspects; first, we extensively utilized several pre-trained networks to evaluate the best architecture for age classification; second, all the previous works fine-tuned their new models based on the features which were extracted from the last convolutional layer of the pre-trained models, while in this work, we proposed to use a dimensionality reduction method in order to enhance the extracted features for this task. Third, our model learns robust features by utilizing several features from distinct domain-based tasks; we combined different features extracted from different pre-trained models into one feature vector to obtain more information about the subject's age from different perspectives, while other works utilized on pre-trained model to find the best features for age estimation.

3. The Adience benchmark

The Adience benchmark is used in this work. The Adience, contains 26 K face images of 2284 subjects who are divided into 8 age groups called labels. Table 1 shows the Adience labels and the number of images per label. Standard five-fold, subject-exclusive cross-validation protocol is applied for dividing the database into a train and test groups. The same settings were used in [1]. The Adience is a challenging database since it consists of unfiltered face images, which were uploaded to the Flicker website using smart phones. The images are not filtered with any manual filtering techniques. Images in the database reflect real-world conditions of uncontrolled environments such as significant variations in pose, expression, lighting, image quality and resolution.

The Adience is not designed for face recognition task so that the number of images per subject is not balanced. Around 80% of the subjects in the database have only one image, while the rests have around 100 to 400 images. When the number of images per subject is small

Table 1
The Adience benchmark.

Gender	Labels (in years)								Total # of im.
	0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–	
F	682	1234	1360	919	2589	1056	433	427	9411
M	745	928	934	734	2308	1294	392	442	8192

for a label while it is bigger for other labels, the classifier will be biased for the labels with more images.

DNNs architectures are susceptible to overfitting due to the large number of parameters required to training such networks. To some degree, augmentation helps to decrease this problem by increasing the number of images per subject to balance the database. In this work, several popular augmentation techniques are applied as follows. Each image is resized to 224 × 224. Each image is flipped horizontally. Each image is rotated using nearest, bilinear, and bicubic interpolation by the angles in {−15, −10, −5, 5, 10, 15}. A Gaussian white noise is added to each image with different mean and variance. Finally, 50 images containing original and augmented images are chosen randomly for each subject in the database.

4. Methodology: utilizing transfer learning for age range classification

This section describes the architectures and network configurations of the CNNs used in this work. Section 4.1 presents the training and estimating the age range on a small database. Section 4.2 presents the training and estimating the age range using a CNN architecture which was trained for face recognition on a large database. Section 4.3 presents the feature dimension reduction using the PCA and its effects on the performance. The Proposed model is shown in Fig.1. The Architectures I and II are designed based on this model.

4.1. Training and estimation on a small database

The CNN architecture [8] is modified and used in this study for age range classification. The modification includes the change in the number of convolution layers and their connections. A small deep network is used to overcome overfitting problem. This is comparable to the state-of-the-art network architecture used in [38], which trains more than 2000 subjects with 1000 different images per person for face recognition. Our architecture includes 8 layers, 5 of which are convolutional layers as filter banks that are used to process the input image. The remainder 3 layers are fully connected layers, where the filters in these layers match the size of the input image. Each layer contains a linear manipulator followed by non-linear manipulators as normalization, rectification, drop-out, and pooling. A stride of 4 is used in the first convolutional layer to keep the computation reasonable and the manipulation process fast for training. Table 2 shows the network architecture in detail.

All the 8 layers are followed by a rectification layer. The first, second, and fifth convolutional layers include non-linear max pooling operator. The first and second convolutional layers are normalized using local response normalization. No pooling or normalization is added to the third and fourth convolutional layers. The first and second fully connected layers contain 4096 neurons; they are regularized with a dropout and normalized with the local response normalization. The last fully connected layer is an N-way classifier with a softmax layer, where N is the number of subjects in the database.

The proposed CNN architecture is trained for face recognition task by using the Adience database. Training RGB images are scaled to 256 × 256 pixel. The resulted scaled images were cropped to 224 × 224 pixel patches, and then the average of all the training images is subtracted. The goal of the training is to maximize the prediction of the softmax layer and to find the optimal network parameters (weights and biases).

The optimization of the N-way classifier is carried out by using stochastic gradient descent algorithm with mini-batches and momentum. The batch size is chosen as 256 and the momentum is set to be 0.9. To avoid overfitting, drop out and weight decay (L2 normalization) are used to regularize the network parameters during the training process. The weight decay is set to 5×10^{-4}. A dropout rate of 0.5 is used after each fully connected layer except the softmax layer. The initial value of the learning rate is set to be 10^{-2}, and then decreased by a factor of 10 whenever there is no improvement for the validation set accuracy. In general, three leaning rates are used to train the network. One of the most critical steps in deep networks is the initialization of the network parameters. In this work, biases are initialized to zero and the weights of the filters are initialized by using the random initialization procedure in [39].

After the CNN is trained for face recognition task, it is modified for age range classification task. The idea here is to train a deep network to study facial features from image and then retrain and fine-tune this network to estimate the age information. The modification of the CNN is performed by removing the fully connected layers and replacing them with four new fully connected layers of different sizes. It is shown in Fig. 1. The sizes of the first three fully connected layers are chosen as 4096, 5000, and 5000, respectively. Each of which is followed by two manipulation layers, one dropout layer, and one normalization layer. The last fully connected layer is a softmax layer with a size of 8, which represents the number of labels in the Adience database. Each label represents an age range. The probability of each label is used to estimate the age of corresponding face image. The eight convolutional layers which were used during the training for face recognition task are reused in the modified CNN as shown in Fig. 1.

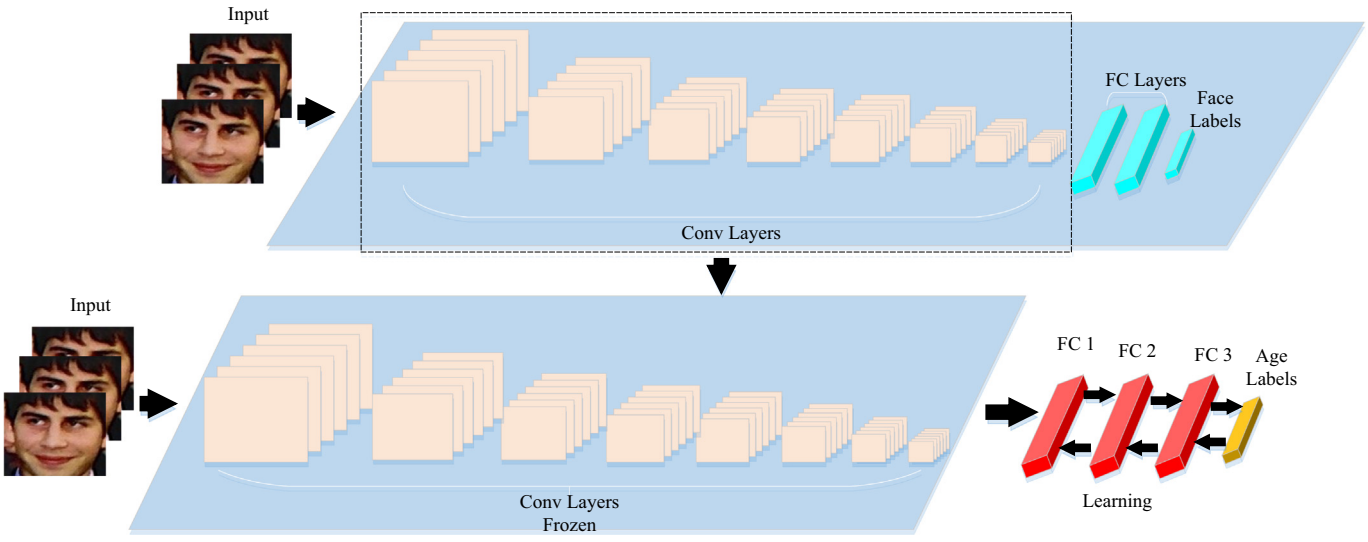

Fig. 1. Proposed model.

Table 2
Architecture I. Network configuration.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13
type	Input	conv	relu	norm	mpool	conv	Relu	norm	mpool	conv	relu	conv	relu	conv
name	n/a	conv1	relu1	norm1	pool1	conv2	relu2	norm2	pool2	conv3	relu3	conv4	relu4	conv5
support	n/a	11	1	1	3	5	1	1	3	3	1	3	1	3
filt dim	n/a	3	n/a	n/a	n/a	64	n/a	n/a	n/a	256	n/a	256	n/a	256
num filts	n/a	64	n/a	n/a	n/a	256	n/a	n/a	n/a	256	n/a	256	n/a	256
stride	n/a	4	1	1	2	1	1	1	2	1	1	1	1	1
pad	n/a	0	0	0	$0 \times 1 \times 0 \times 1$	2	0	0	0	1	0	1	0	1

layer	14	15	16	17	18	19	20	21	22	23	24	25	26
type	Relu	mpool	conv	relu	dropout	conv	Relu	dropout	conv	relu	dropout	conv	softmxl
name	relu5	pool5	fc6	relu6	dropout6	fc7	relu7	dropout7	fc8	relu8	dropout8	fc8	loss
support	1	3	6	1	1	1	1	1	1	20	21	1	1
filt dim	n/a	n/a	256	n/a	n/a	4096	n/a	n/a	5000	n/a	n/a	5000	n/a
num filts	n/a	n/a	4096	n/a	n/a	5000	n/a	n/a	5000	n/a	n/a	1743	n/a
stride	1	2	1	1	1	1	1	1	1	1	1	1	1
pad	0	0	0	0	0	0	0	0	0	0	0	0	0

For each convolutional layer, number of filters, the filter size, their receptive field convolution stride, and spatial padding are indicated.

The stochastic gradient descent algorithm is used to train the modified CNN in order to find the optimal parameters which will enable the network to achieve better classification accuracies. Images scaled 224×224 pixels are used. The output of each layer is forwarded to the next layer as an input until the softmax layer calculates the probability of each label. The learning is performed only on the fully connected layers. It means that the parameters of the convolutional layers are frozen, while the parameters of the fully connected layers are allowed to be changed. It is shown in Fig. 1. Freezing the training on the convolutional layers ensures that the process of extracting facial features is unchanged. The learning rate is set initially to 0.1 and then decreased by a factor of 10 if there is no improvement in the validation set learning. The dropout value is chosen as 0.6.

It is observed that using a weight decay together with dropout technique have a positive effect on the classification accuracies [40]. The weight decay is normally set to 10^{-4} or 10^{-5}. The later value has worked fine for the modified network, but increasing the value to 10^{-3} has provided a higher overall accuracy. It is being said, deploying dropout technique together with the weight decay as regularizers has no negative effects on the training of the fully connected layers. On the contrary, increasing the value of the weight decay to 10^{-3} enhances the learning process and the accuracies. Since the convolutional layers are trained previously, and training occurs only on the newly added fully connected layers, we are able to increase the weight decay value with an increased network performance. The modified network has faster convergence as it only trains the fully connected layers as oppose to train all layers in the original CNN.

A given test image is rescaled to 256×256, and then three images of 224×224 pixels are extracted. The first image is extracted from the center of the original image. The second and the third images are cropped from the upper-right and the bottom-left corners of the original image, respectively. Then, the trained network is applied densely on the three images. The softmax probability score vectors of the three images are averaged to obtain a final vector of class scores for the original test image from the three images. This method reduces the impact of the challenges such as low-resolutions, various expressions, and occlusions in the database.

4.2. Training on a large database based on a trained deep face recognition model

As mentioned earlier, using an efficient facial feature extractor is expected to perform well for age range classification. However, a trained face recognition model that was trained on a large database may extract facial features more efficiently than training a new model on a small database. The proposed architecture employs a previously trained model for face recognition and performs age range classification on the Adience. The idea here is to take advantage of a large database and the deep architecture of the network, which is designed on a large database. Deep network architectures trained on large databases are capable to extract distinctive and robust features, and they are less prone to overfitting. The available databases for age range classification such as the Adience benchmark are relatively small compared to the databases for face recognition. Building deep network architecture for age range classification or face recognition on small databases is expected to have poor performance.

In this work, we consider the CNN architecture proposed by [38]. It achieved comparable results to the state-of-the-art for face recognition task (VGG-Face). In [38], three CNN architectures named A, B, and C were used. We use the architecture A. Details of the large database and the configuration of the architecture A can be found in [38]. The architecture A consists of eight convolutional layers and three fully connected layers. A rectification operator is used after each convolutional operator. A max pool operator is added at the end of each convolutional layer. A 4096 dimensional output is used for the first two fully connected layers. Dropout with $p = 0.5$ and rectification operator are applied to the first two layers. N-way class prediction is used for optimizing the network parameters. Therefore, the output size of the last layer is chosen to be 2622, which represents the number of subjects in the large database.

Here, the convolutional layers of the CNN VGG-Face are reused, while the fully connected layers are replaced with four new layers. The details of the proposed network architecture and configuration are given in Table 3. The weights between the new layers are initialized by a Gaussian distribution with zero mean and 10^{-2} standard deviation. The new network is trained only for the newly added fully connected layers while keeping the original convolutional layers frozen during the training. This approach appears to be very fast since only the newly added fully connected layers are trained.

The same prediction method given earlier in section 4.1 is used.

4.3. Reducing the dimension of the extracted features

There are many cases where the measured or the observed data vectors are described as a high dimensional data. Normally, a significant portion of the high dimensional data is redundant and has low variance, undesired, or resulted from linear operations over other desired data. The goal of dimensionality reduction is to reduce the dimension of the

Table 3
Architecture II. Network configuration.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
type	Input	Conv	relu	norm	mpool	conv	Relu	Norm	mpool	conv	Relu	Conv	relu	conv	Relu
name	n/a	conv1	relu1	norm1	pool1	conv2	relu2	norm2	pool2	conv3	relu3	conv4	relu4	conv5	relu3_2
support	n/a	3	1	3	1	2	3	1	3	1	2	3	1	3	1
filt dim	n/a	3	n/a	64	n/a	n/a	64	n/a	128	n/a	n/a	128	n/a	256	n/a
num filts	n/a	64	n/a	64	n/a	n/a	128	n/a	128	n/a	n/a	256	n/a	256	n/a
stride	n/a	1	1	1	1	2	1	1	1	1	2	1	1	1	1
pad	n/a	1	0	1	0	0	1	0	1	0	0	1	0	1	0

layer	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
type	conv	Relu	mpool	conv	relu	conv	Relu	Conv	relu	mpool	Conv	Relu	conv	relu	Conv
name	conv3_3	relu3_3	pool3	conv4_1	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3
support	3	1	2	3	1	3	1	3	1	2	3	1	3	1	3
filt dim	256	n/a	n/a	256	n/a	512	n/a	512	n/a	n/a	512	n/a	512	n/a	512
num filts	256	n/a	n/a	512	n/a	512	n/a	512	n/a	n/a	512	n/a	512	n/a	512
stride	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1
pad	1	0	0	1	0	1	0	1	0	0	1	0	1	0	1

layer	30	31	32	33	34	35	36	37	38	39	40	41	42
type	relu	Mpool	conv	relu	dropout	conv	Relu	dropout	conv	relu	Dropout	Conv	softmx
name	relu5_3	pool5	fc6	relu6	drop6	fc7	relu7	drop7	fc8	relu8	drop8	fc8	prob
support	1	2	7	1	1	1	1	1	1	1	1	1	1
filt dim	n/a	n/a	512	n/a	n/a	4096	n/a	n/a	5000	n/a	n/a	5000	n/a
num filts	n/a	n/a	4096	n/a	n/a	5000	n/a	n/a	5000	n/a	n/a	8	n/a
stride	1	2	1	1	1	1	1	1	1	1	1	1	1
pad	0	0	0	0	0	0	0	0	0	0	0	0	0

For each convolutional layer, number of filters, the filter size, their receptive field convolution stride, and spatial padding are indicated.

high dimensional data to a smaller one while preserving the same useful or desired information. There are many benefits of dimensionality reduction, for example, it reduces the space needed to store the data during training, it decreases the time needed to process the data, and it increases the performance of the data in many classification tasks [41, 42]. In age and face recognition tasks, the number and the size of the input images are considerably large and require careful processing in order to extract and select the distinctive features. As shown in Fig. 2, different deep convolutional models that were trained for different tasks other than age range classification are used. PCA is applied to the last convolutional layer output of these models for dimensionality reduction.

The input images are fed to each deep trained model until the output features of the last convolutional layer are calculated. These features are stacked together for the entire training data set. This results in a high dimensional feature vector. The size of the last convolutional layer differs between the trained models, and it is large for all models. As a result, dimensionality reduction is required to fine tune the feature vectors. Since each trained model was trained for a different task, it is expected to have different feature vectors with a different level of performance in age range classification. In this work, the dimensionally reduced features from each trained model, and their combination are examined. PCA technique is used for dimensionality reduction. PCA transforms the large space into a smaller subspace using linear transformation.

5. Results and discussions

Several experiments have been conducted to evaluate the performance and efficiency of the proposed models. The experiments are carried out on a Titan X with 3072 cores and 12 GB of video memory. Exact and 1-off accuracies for eight age labels and overall accuracy are calculated.

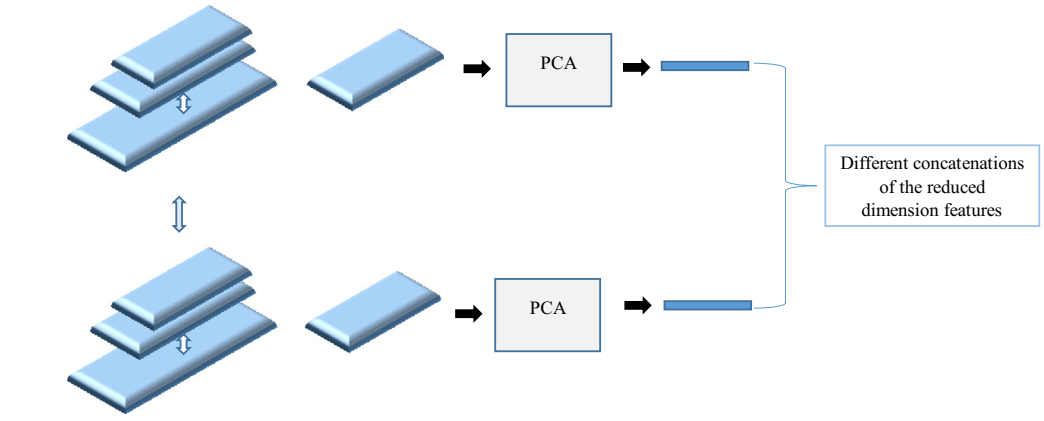

Fig. 2. Dimensionality reduction.

5.1. Results without dimensionality reduction

43.0% overall accuracy is achieved using the architecture I without augmentation. The overall accuracy of the architecture I is improved to 45.83% by augmentation since the number of training images per subject is increased by the augmentation, causing an improvement in the accuracy. The highest accuracy (59.90%) is achieved by the architecture II, which is based on the VGG-Face (Table 4). This improvement supports the main idea of this work. The employment and retraining of a very deep and well-trained CNN for face recognition improve the performance of age range classification. Confusion matrices of two architectures are given in Tables 5 and 6.

As it can be seen from Tables 5 and 6, the classification accuracy for some classes was significantly better than other classes and the highest misclassification ratios occur between adjacent classes. However, from Table 5, we can consider that the highest ratios of misclassification for most of the classes occur with class 25–32, this might be due to the fact that the number of samples for this class is significantly larger than the rest of the classes in the Adience database. Comparing the results in Table 6 with Table 5, it can be considered that the performance accuracy of the classes is enhanced, but we still consider that some classes still do poorly. The proposed architecture (Architecture II) decreased the negative influence of the imbalanced classes especially when the for non-adjacent classes, this can be noticed in class 43–58 and class 60–. For the adjacent classes to class 25–32, some enhancement can be noticed, but still there is a highly misclassification occurred with this class.

To emphasize the effectiveness of the Architecture II for age range classification, we evaluated five more architectures for further evaluation: The GoogLeNet [43] and ResNet-50 [44] architectures which performed exceptionally well in ImageNet ILSVRC14; the VGG-VD [45] models with 16 and 19 layers, trained on ImageNet ILSVRC for image classification; and FNC-8s [46] trained for semantic segmentation. These are retrained, fine-tuned and tested for age classification. The architecture of the fully connected layers used for each CNN is summarized in Table 7. We architect and fine-tune these CNNs to predict the age by changing the fully connected layers and their number nodes. Then, each network is trained and fine-tuned while the convolutional layers are kept frozen during the training as explained in Section 4.1. The accuracy results of these CNNs for age classification are presented in Table 8.

After fine-tuning each network from their original task to age classification task, GoogLeNet, ResNet-50, VGG-VD-16, VGG-VD-19, and FNC-8s achieved 45.07%, 42.46%, 45.01%, 45.99%, and 43.87% of accuracy, respectively. For all networks, it is observed that the best results are achieved after convergence at 3 different learning rates as 0.01, 0.001, and 0.0001. Fine-tuning from different classification tasks to age classification provides reasonable accuracies compared to the state-of-art results.

Table 4
Accuracy performance (%).

	Architecture II		Architecture I with augmentation		Architecture I without augmentation	
	Acc.					
Labels	Exact	1-off	Exact	1-off	Exact	1-off
0–2	93.17	99.59	80.54	93.15	79.30	89.65
4–6	62.11	96.32	41.23	83.16	29.47	71.93
8–13	42.06	61.18	37.94	62.65	29.12	44.71
15–20	24.23	93.39	11.01	86.79	13.22	85.02
25–32	86.17	92.71	70.08	81.82	69.51	84.56
38–43	8.88	91.52	14.79	72.58	13.61	72.19
48–53	38.17	80.50	10.79	20.75	11.20	28.63
60–	60.70	95.72	26.46	27.63	28.79	29.57
Overa. Acc.	**59.90**	**90.57**	**45.83**	**72.99**	**43.00**	**70.42**

Table 5
Confusion matrix for the architecture I with augmentation.

		Act.							
		0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–
Pred.	0–2	**80.54**	12.61	3.11	0.00	2.90	0.41	0.02	0.41
	4–6	22.46	**41.23**	19.47	5.09	10.88	0.00	0.35	0.53
	8–13	2.65	13.53	**37.94**	11.18	32.35	0.59	0.59	1.18
	15–20	1.76	3.08	18.50	**11.01**	57.27	5.29	1.32	1.76
	25–32	1.23	2.65	13.26	6.91	**70.08**	4.83	0.85	0.19
	38–43	2.17	3.35	10.45	4.14	56.61	**14.79**	1.18	7.30
	48–53	3.32	0.00	9.13	8.30	58.50	8.71	**10.79**	1.24
	60–	3.11	0.00	14.79	3.50	36.58	14.40	1.17	**26.46**

From Table 8, it is observed that networks previously trained for image classification task except ResNet-50 perform better than the FNC-8s, which was pre-trained for semantic segmentation. It can be stated that CNNs, which were pre-trained for image classification can contain more age related features compared to CNNs pre-trained for image semantic segmentation. The classification accuracy of some classes was significantly better than that of the other classes, and the highest misclassification ratios occurred between adjacent classes. The confusion matrices revealed that the highest misclassification ratios for most of the classes occurred with the class 25–32. This might be due to the fact that the number of samples for this class is significantly larger than the rest of the classes in the Adience database. Although ResNet-50 was trained for image classification task, its performance, in terms of overall accuracy, is slightly lower than the other CNNs that were trained for the same task. Moreover, the accuracy results of ResNet-50 for 8–13, 38–43, 48–53, 60– age groups are zero. This might be because of the very deep architecture of the ResNet-50 network, which fine-tuned on a relatively small database (Adience), and this resulted in a biased training toward the classes with large number of samples. Hence, to achieve better results while utilizing transfer learning with such deep architectures (such as ResNet-50), the database size should be sufficiently large, and the number of training samples should be semi balanced between the classes.

These results support the fact that both the number of the training images and subjects of the used database and the pre-training task of the CNN determine the network ability to achieve good results for age classification from facial images.

5.2. Results with dimensionality reduction

Different experiments have been conducted to evaluate the effectiveness of the dimensionally reduced features for age range classification based on different trained models for different tasks. The accuracy results of different CNN models and their combinations for age range classification are given in Table 9 accompanied by the DNN specifications used for each experiment. Table 10 provides the confusion matrix for the model that achieved the highest accuracies.

Since the optimal configuration settings for CNNs are problem dependent, different settings have been tested to reach the best results for age range classification from facial images. Four fully connected

Table 6
Confusion matrix for the architecture II.

		Act.							
		0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–
Pred.	0–2	**93.17**	6.42	0.21	0.00	0.21	0.00	0.00	0.00
	4–6	26.84	**62.11**	7.37	1.93	1.58	0.00	0.18	0.00
	8–13	1.76	6.18	**42.06**	12.94	35.59	0.29	1.18	0.00
	15–20	1.76	0.44	4.41	**24.23**	64.76	0.00	4.41	0.00
	25–32	0.00	0.09	0.85	3.22	**86.17**	3.31	4.83	1.52
	38–43	0.39	0.20	0.39	1.78	59.76	**8.88**	22.88	5.72
	48–53	0.00	0.00	0.00	0.00	19.50	8.71	**38.17**	33.61
	60–	0.00	0.00	0.00	1.17	1.95	1.17	35.02	**60.70**

Table 7
Different CNN architectures.

Network	# of fully connected layers	# of nodes/layer
GoogLeNet	4	1024, 2048, 2048, 8
ResNet-50	4	2048, 5000, 5000, 8
VGG-VD-16	4	4096, 6000, 6000, 8
VGG-VD-19	4	4096, 6000, 6000, 8
FNC-8s	4	4096, 5000, 5000, 8

layers for all models are tested to be the best for achieving the highest accuracy (Table 9). The number of nodes in each layer is different between the layers and between different trained models. For some models, the number of nodes in the first fully connected layer is greater than the number of nodes in the rest of the fully connected layers, while for other models, the number of nodes in the input fully connected layer is smaller than the rest of the fully connected. For the FNC-8s model, the number of nodes in each fully connected layer is the same. Optimum dropout and weight decay values are chosen after extensive experiments. For the GoogLeNet and FNC-8s models, the dropout rate value is chosen as 0.5, while it is set to 0.8 for the other models. For the weight decay value, we tested different values, ranged from 10^{-5} to 10^{-2}. Most of the models performed well at 10^{-3} while for the GoogLeNet achieved the best performance at 10^{-4}.

One of the main steps for performing a successful classification task is to select and extract the most relevant features to problem under study. For simple problems, this task is straight forward, but for harder problems where the feature space is complex, complex representations are required to extract the relevant features. In this work, the problem is to estimate the subject's age from his\her facial image; this is a multiclass problem; the type of images is unconstrained, and the classes are not balanced. The proposed usage of deep CNNs is to represent the complex feature space and later to extract the effective features related to subject's age. Although the deep CNNs were successful to extract age related features, the dimension of this extracted features was relatively high, which means it contains age related features but with different degrees of relevance for age range classification. The dimensionality reduction step helps to select the most relevant features which improves the achieved results. As well as, reducing the dimensionality of the features obtained from the last convolutional layers of each model helps in combining all the features from these models. In this way, each image is represented with multiple feature sets obtained from all models, instead of using one feature set. Hence, these distinctive reduced feature sets enhance the performance accuracy of age range classification task.

In general, it is observed from the Tables 8 and 9 that the dimensionality reduction improves the classification performance of all CNNs models for the age classification task and decreases the negative influence of the imbalanced classes especially for non-adjacent classes. It is also observed that the best results are achieved with the VGG-Face CNN model. Another observation is the significant improvement by the FNC-8s with dimensionality reduction. FNC-8s achieved the highest accuracy among the other networks except the VGG-Face. Moreover,

Table 8
Overall accuracies of different CNN architectures (%).

Label	GoogLeNet	ResNet-50	VGG-VD-16	VGG-VD-19	FNC-8s
0–2	86.75	90.89	84.27	83.44	79.92
4–6	27.89	15.79	42.28	43.68	30.35
8–13	21.47	0.29	33.82	30.59	29.71
15–20	14.10	0.00	14.98	11.89	18.50
25–32	76.61	97.82	58.90	62.03	66.29
38–43	12.03	0.00	24.06	27.42	17.36
48–53	7.05	0.00	12.45	10.79	4.98
60–	34.63	0.00	33.46	35.02	43.97
Overall Acc.	**45.07**	**42.46**	**45.01**	**45.99**	**43.87**

Table 9
Overall accuracies of different CNN models with dimensionality reduction.

Trained model	Accuracy %	# of fully connected layers	# of nodes in each layer	Dropout rate	Weight decay
VGG-Face	60.60	4	512-1024-1024	0.8	10^{-3}
GoogLeNet	46.43	4	512-1024-1024	0.5	10^{-4}
ResNet-50	45.69	4	512-1024-1024	0.5	10^{-4}
VGG-VD-16	47.13	4	512-1024-1024	0.8	10^{-3}
VGG-VD-19	47.49	4	512-1024-1024	0.8	10^{-3}
FNC-8s	48.95	4	512-512-512	0.5	10^{-3}
VGG-Face + FNC-8s	61.39	4	1024-512-512	0.8	10^{-3}
VGG-VD-19 + FNC-8s	51.88	4	1024-512-512	0.8	10^{-3}
Combined Models (VGG-Face + GoogLeNet + ResNet-50 + VGG-VD-16 + VGG-VD-19 + FNC-8s)	62.26	4	3072-1024-1024	0.8	10^{-3}

The network settings for each experiment are indicated.

the performances get better when all features of different networks are combined together. For example, combining the features of the FNC-8s with the features of the VGG-VD-19 achieved better results than their individual use. The best results are achieved when all networks are combined with dimensionality reduction. Our results support the validity of our proposed idea: A previously trained CNNs for face recognition on a large database can be effectively retrained and fine-tuned to design a DNN for age range classification on another database.

Fig. 3 shows some of the challenging images from the Adience database. These images are classified correctly by using the proposed work although they consist of image formation distortions such as motion blur, low-resolution, pose, and facial expressions.

5.3. Comparison with previous works

The proposed architectures and models are compared with state-of-the-art results in Table 11. As a general notice, face recognition-based age range classification model outperforms state-of-art results. On the other hand, the VGG-VD-19 + FNC-8s model with dimensionality reduction outperforms state-of-art even though VGG-VD-19 and FNC-8s models were not previously trained for face recognition.

Building an efficient large database containing millions of face images for age classification is a difficult task due to the fact that this requires an access to participants' private information and requires IRB approval to do so. Furthermore, the collected images need to be manually labeled. Therefore, databases that are designed for age classification from dynamic and real environments such as social sites are limited on their size. They are not comparable in size with other databases which have been designed for object recognition or face recognition such as ImageNet [47] and Pascal [48] databases.

Table 10
Confusion matrix for the combined models with dimensionality reduction.

		Act.							
		0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–
Pred.	0–2	**86.54**	12.84	0.41	0.00	0.21	0.00	0.00	0.00
	4–6	25.09	**62.28**	10.18	1.05	0.70	0.18	0.00	0.53
	8–13	0.59	7.06	**45.88**	22.65	20.88	2.35	0.00	0.59
	15–20	0.00	0.44	8.37	**31.28**	56.39	3.52	0.00	0.00
	25–32	0.19	0.19	2.37	3.13	**77.37**	16.29	0.28	0.19
	38–43	0.00	0.20	0.59	2.76	35.70	**47.14**	3.55	10.06
	48–53	0.00	0.00	0.83	0.00	9.54	55.60	**11.62**	22.41
	60–	0.00	0.00	0.00	0.00	2.33	10.89	7.00	**79.77**

Fig. 3. Some of the challenging images classified correctly by this work.

The work in [6] was the first step in age and gender classification from face images by using DNNs. It employed a relatively simple and shallow network for feature extraction and classification stages and proposed an over-sampling technique to solve the misalignment challenge partially. Zhu et al. [49] studied a DCNN for a multitask learning to train shared features for age and gender tasks in end-to-end manner. A cascaded CNN is introduced in age estimation in [30]. It consists of three modules: age group classifier, DCNN bases regressors, and errorenous age prediction. Rothe et al. [36] utilized a deep CNN architecture based on the VGG-16 model that was pre-trained for image classification. It does not rely on facial landmarks to extract facial age related features. The pretrained DNN was used as a facial feature extractor. In [30,36], the models are mainly built and trained for apparent age estimation. Since apparent age estimation and real age estimation are different tasks, the reported improvements are limited when compared with other works. A new face descriptor model is presented in [50] based on three attributes: 1) The age primitives which finds the crucial texture primitives; 2) The latent second direction to keep the structural information; 3) The global adaptive threshold to discriminate in the flat and textured region. The new descriptor was used to extract facial features for age classification. In Liu et al.'s work [51], the label-sensitive deep metric learning (LSDML) is proposed to learn a discriminative feature similarity in facial age estimation. The goal was to exploit the label correlation between the training face samples in term of the labels (subspace) to achieve balanced training samples.

Table 11
Comparison of state of art results (%).

Method		Exact accuracy	1-off accuracy
[1]	LBP	41.4	78.2
	LBP + FPLBP	44.5	80.7
	LBP + FPLBP + Dropout 0.5	44.5	80.6
	LBP + FPLBP + Dropout 0.8	45.1	79.5
[6]	Shallow CNNs Using Single Crop	49.5	84.6
	Shallow CNNs Using Over-Sample	50.7	84.7
[49]	Employ light weight DCNN for a multitask learning scheme (age + gender), best model single-6-conv	49.7	–
[30]	Cascaded Convolutional Neural Network	52.88	88.45
[36]	DCNNs based on VGG-16 architecture + softmax expected function for refinement	55.6	89.7
[50]	Subject-Exclusive DAPP	54.9	–
	Subject-Inclusive DAPP	62.2	–
[51]	LSDML: w/o data augmentation	56	–
	LSDML: random cropping + horizontal flipping	56.9	–
	M-LSDML: w/o data augmentation + 3 DB	58.2	–
	M- LSDML: random cropping + horizontal flipping + 3 DB	60.2	–
This work	Architecture 1 with augmentation	45.83	72.99
	Architecture 2	59.9	90.57
	Architecture 2 with dimensionality reduction	60.60	92.25
	VGG-Face + FNC-8s with dimensionality reduction	61.39	92.31
	VGG-VD-19 + FNC-8s with dimensionality reduction	51.88	82.72
	Combined-Models with dimensionality reduction	62.26	92.63

From Table 11, it can be considered that, the result in [50] is close to the results in this work. However, this result cannot be compared to results of this work because in [50] the subject-inclusive protocol is used, which allows same subjects to appear in test and train samples. While the proposed methods in this work use the subject-exclusive protocol, which does not allow the same subjects to appear in train and test samples.

Overfitting is one of the biggest challenges of machine learning, especially when using small databases. It becomes an even more common and significant issue in DNNs, where the networks often have a large

number of layers containing thousands of neurons. Hence, the number of connections in these networks is astronomical, reaching to the millions. As a result, a compact architecture network should be designed to trade-off between overfitting and network complexity. If the network is not complex enough, it may not be powerful to capture the necessary information to gain more accurate result. In this work, we take advantage of the large databases that are originally formed and used for face recognition in order to estimate the age information from face images.

6. Conclusions

Age range classification of the subjects from their face images is considered as an important task for many applications. Although its importance is recognized, it has received less attention than other image classification tasks. Unlike most of the previous studies that used constrained face images, we used a database that has unconstrained face images reflecting the variations of the real subject images taken from the internet repositories. The main goal of the work is to investigate the employment of CNNs, which were previously trained for different tasks on large databases, in the design of a DNN for age range classification task. We have proposed and evaluated several CNN architectures.

A deep pre-trained CNN model for face recognition is used to extract facial features. Then, these extracted facial features are used to train a DNN for age range classification. In addition, dimensionality reduction is performed on the last convolutional layer features of previously pre-trained CNN models on different tasks other than age range classification. The dimensionally reduced features are then incorporated and trained to estimate the age by using DNN architecture. We provide extensive experimental results, which demonstrate the capability of our proposed work to classify the age from the facial images. The proposed work significantly outperformed stat-of-the-art results by approximately 12% on the Adience benchmark.

Despite the difficulty of building a large unconstrained real-word benchmark containing millions of face images from social media websites and internet repositories for age range classification, we hope that such benchmarks will become available in near future. The availability of large benchmarks will help improving the current results further especially by using very deep CNNs that have shown remarkable performances in other classification fields.

Finally, we can conclude that the idea of our proposed work can be used in other classification fields with relatively small databases. Deep networks can be designed by using pre-trained CNNs by using large databases from related domain specific tasks to overcome the challenges of relatively small databases.

Acknowledgement

No conflict of interest.

References

[1] E. Eidinger, R. Enbar, T. Hassner, Age and gender estimation of unfiltered faces, IEEE Transactions on Information Forensics and Security 9 (12) (2014) 2170–2179.
[2] Y. Fu, G. Guo, T.S. Huang, Age synthesis and estimation via faces: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 1955–1976.
[3] A.C. Gallagher, T. Chen, Understanding images of groups of people, IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2009, pp. 256–263.
[4] H. Han, C. Otto, A.K. Jain, Age estimation from face images: human vs. machine performance, in International Conference on Biometrics (ICB) (2013) 1–8.
[5] W.-L. Chao, J.-Z. Liu, J.-J. Ding, Facial age estimation based on label-sensitive learning and age-oriented regression, Pattern Recogn. 46 (3) (2013) 628–641.
[6] G. Levi, T. Hassner, Age and gender classification using convolutional neural networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2015) 34–42.
[7] H. Han, C. Otto, X. Liu, A.K. Jain, Demographic estimation from face images: human vs. machine performance, IEEE Trans. Pattern Anal. Mach. Intell. 37 (6) (2015) 1148–1161.
[8] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems 2012, pp. 1097–1105.
[9] Y.H. Kwon, N. da Vitoria Lobo, Age classification from facial images, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR'94 1994, pp. 762–767.
[10] L.G. Farkas, Anthropometry of the Head and Face, Raven Press, New York, 1994.
[11] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, H. Koshimizu, Age and gender estimation from facial image processing, Proceedings of the 41st SICE Annual Conference 2002, pp. 13–18.
[12] J. Hayashi, M. Yasumoto, H. Ito, H. Koshimizu, Method for estimating and modeling age and gender using facial image processing, in Proceedings of Seventh International Conference on Virtual Systems and Multimedia (2001) 439–448.
[13] N. Ramanathan, R. Chellappa, Modeling age progression in young faces, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) 2006, pp. 387–394.
[14] A. Lanitis, C.J. Taylor, T.F. Cootes, Toward automatic simulation of aging effects on face images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 442–455.
[15] X. Geng, Z.-H. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns, IEEE Trans. Pattern Anal. Mach. Intell. 29 (12) (2007) 2234–2240.
[16] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, H. Dai, Learning from facial aging patterns for automatic age estimation, Proceedings of the 14th ACM International Conference on Multimedia 2006, pp. 307–316.
[17] Y. Fu, Y. Xu, T.S. Huang, Estimating human age by manifold analysis of face pictures and regression on aging features, in IEEE International Conference on Multimedia and Expo (2007) 1383–1386.
[18] K. Scherbaum, M. Sunkel, H.P. Seidel, V. Blanz, Prediction of individual non-linear aging trajectories of faces, Computer Graphics Forum 2007, pp. 285–294.
[19] Y. Fu, T.S. Huang, Human age estimation with regression on discriminative aging manifold, IEEE Transactions on Multimedia 10 (4) (2008) 578–584.
[20] A. Gunay, V.V. Nabiyev, Automatic age classification with LBP, 23rd International Symposium on Computer and Information Sciences, ISCIS'08 2008, pp. 1–4.
[21] F. Gao, H. Ai, Face age classification on consumer images with gabor feature and fuzzy lda method, International Conference on Biometrics, pp. 132-141, 2009.
[22] S. Yan, M. Liu, T.S. Huang, Extracting age information from local spatially flexible patches, in IEEE International Conference on Acoustics, Speech and Signal Processing (2008) 737–740.
[23] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, T.S. Huang, Regression from patch-kernel, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2008) 1–8.
[24] J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen, W. Gao, Design sparse features for age estimation using hierarchical face model, 8th IEEE International Conference on Automatic Face & Gesture Recognition, FG'08 2008, pp. 1–6.
[25] G. Mu, G. Guo, Y. Fu, T.S. Huang, Human age estimation using bio-inspired features, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2009) 112–119.
[26] G. Guo, G. Mu, Y. Fu, C.R. Dyer, T.S. Huang, A study on automatic age estimation using a large database, in ICCV (2009) 1986–1991.
[27] C. Shan, Learning local features for age estimation on real-life faces, Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis 2010, pp. 23–28.
[28] F. Alnajar, C. Shan, T. Gevers, J.-M. Geusebroek, Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions, Image Vis. Comput. 30 (12) (2012) 946–953.
[29] R. Ranjan, S. Zhou, J. Cheng Chen, A. Kumar, A. Alavi, V.M. Patel, et al., Unconstrained age estimation with deep convolutional neural networks, Proc. IEEE Int. Conf. Comput. Vis. Workshop (2015) 109–117.
[30] J.-C. Chen, A. Kumar, R. Ranjan, V.M. Patel, A. Alavi, R. Chellappa, A cascaded convolutional neural network for age estimation of unconstrained faces, Proc. IEEE Int. Conf. Biometr. Theory Applicat. Syst. (2016) 1–8.
[31] H.-F. Yang, B.-Y. Lin, K.-Y. Chang, C.-S. Chen, Automatic age estimation from face images via deep ranking, Proc. British Mach. Vis. Conf 2015, p. 55.
[32] D. Yi, Z. Lei, S.Z. Li, Age estimation by multi-scale convolutional network, Proc. Asian Conf. Comput. Vis 2014, pp. 144–158.
[33] X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, X. Chen, Agenet: deeply learned regressor and classifier for robust apparent age estimation, Proc. IEEE Int. Conf. Comput. Vis. Workshop (2015) 258–266.
[34] Z. Qawaqneh, A.A. Mallouh, B.D. Barkana, Age and gender classification from speech and face images by jointly fine-tuned deep neural networks, Expert Syst. Applicat. 85 (Nov. 2017) 76–86.
[35] S. Escalera, J. Fabian, P. Pardo, X. Baró, J. Gonzalez, H.J. Escalante, et al., Chalearn looking at people 2015: apparent age and cultural event recognition datasets and results, in Proceedings of the IEEE International Conference on Computer Vision Workshops (2015) 1–9.
[36] R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, Int. J. Comput. Vis. 126 (Aug. 2018) 144–157.
[37] R. Ranjan, S. Sankaranarayanan, C.D. Castillo, R. Chellappa, An all-in-one convolutional neural network for face analysis, in Proceedings of the IEEE International Confernce on Automatic Face & Gesture Recognition (2017) 17–24.
[38] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, British Machine Vision Conference 2015, p. 6.
[39] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, Aistats 2010, pp. 249–256.
[40] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[41] L. Cao, K.S. Chua, W. Chong, H. Lee, Q. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, Neurocomputing 55 (1) (2003) 321–336.
[42] A. Ghodsi, Dimensionality reduction a short tutorial, Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada 37 (2006) 38.
[43] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 1–9.
[44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) 770–778.
[45] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, ICLR, 2015.
[46] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015) 3431–3440.
[47] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[48] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[49] L. Zhu, K. Wang, L. Lin, L. Zhang, Learning a lightweight deep convolutional network for joint age and gender recognition, Proc. IEEE Int. Conf. Pattern Recog. (2016) 3282–3287.
[50] M.T.B. Iqbal, M. Shoyaib, B. Ryu, M. Abdullah-Al-Wadud, O. Chae, Directional age-primitive pattern (DAPP) for human age group recognition and age estimation, IEEE Trans. Inf. Forensics Security 12 (11) (Nov. 2017) 2505–2517.
[51] H. Liu, J. Lu, J. Feng, J. Zhou, Label-sensitive deep metric learning for facial age estimation, IEEE Trans. Inf. Forensics Security 3 (2) (Feb. 2018) 292–305.