Machine Vision and Applications (2019) 30:177–187
https://doi.org/10.1007/s00138-018-0976-1

SHORT PAPER

# Age estimation in facial images through transfer learning

F. Dornaika[1,2] · I. Arganda-Carreras[1,2,3] · C. Belver[1]

Received: 14 March 2018 / Revised: 2 September 2018 / Accepted: 8 September 2018 / Published online: 20 September 2018

**Abstract**

This paper was aimed to address the problem of image-based human age estimation. It has the following main contributions. First, we provide a comparison of three hand-crafted image features and five deep convolutional neural networks (DCNNs). Secondly, we show that the use of pre-trained DCNNs as feature extractors can transfer the knowledge of DCNNs to new datasets and domains that were not necessarily addressed in the training phase. This is achieved by only retraining a shallow regressor over the deep features. Thirdly, we provide a cross-database evaluation involving biological and apparent ages. The paper shows that transfer learning allows the use of pre-trained DCNNs regardless of the type of ages (apparent or biological) that is adopted in DCNN training. The experiments are carried out on three public databases: MORPH, PAL, and Chalearn2016.

**Keywords** Age estimation · Deep CNNs · Transfer learning · Regression · Cross-database evaluation · Mean Absolute Error

## 1 Introduction

The increasing interest in video-based security systems and social robotics has boosted research on the numerical analysis of human faces. Thus, face recognition, face detection, gender classification, and facial expression recognition have attracted much attention in computer vision and pattern recognition field [23,27,28,34,43]. Estimating the age of a subject from the analysis of his/her face image is a relatively new research topic. Age estimation by numerical analysis of the face image has several innovative applications such as intelligent human–machine interfaces and improving protection of minors in various and diverse environments (transport, leisure, medicine, etc.). It can be very useful for demographic statistics collection, intelligent video surveillance, customer profiling, and mining optimization in large databases. The age information could also be used for enriching the tools used in police investigations. In general, automatic age estimation by a computer is useful in applications where the goal is to estimate the age of a person without identifying him. Estimating the age from a face image can use a machine learning method to learn a model for extracted features and then estimate the age for query faces based on the trained model. Age estimation can be viewed as a regression problem, a multi-class classification problem, or a composite of these two. The anthropometry-based method essentially relies on distances of different facial points. The earliest work on image-based age classification was presented in [26]. The proposed approach relies on computing ratios in order to distinguish babies from others. In [15], the authors described a neural network to locate facial points and compute several geometric ratios and distances which are exploited for determining the age from a facial image. The anthropometry-based methods might be useful for distinguishing babies, children, and young adults, but they are useless for estimating the age of adults since they discard the facial skin appearance that is the main source of information about age, ethnicity, and gender. Predicting the human age from a facial image may require a great amount of information from the input face image. Thus, image descriptors are very important since the performance of the age estimation system depends on the quality of the extracted features. Many researches on age estimation have been carried out in order to extract aging features

✉ F. Dornaika
fadi.dornaika@ehu.es

I. Arganda-Carreras
ignacio.arganda@ehu.eus

C. Belver
cbelver001@ikasle.ehu.es

[1] University of the Basque Country UPV/EHU, Manuel Lardizabal, 1, 20018 San Sebastián, Spain

[2] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

[3] DIPC, San Sebastián, Spain

from images. Examples of these features are: active appearance model (AAM) [7], age manifold [12], AGing pattern Subspace (AGES) [14], and biologically inspired features (BIF) [20]. Image-based age prediction methods consider the face image as a texture pattern. Many texture features have been used in order to predict the demographic attributes (age, gender, and ethnicity) like local binary patterns (LBPs) [3], histograms of oriented gradients (HOG) [8], BIF, binarized statistical image features (BSIF) [24] and local-phase quantization (LPQs). BIF and its variants are widely exploited in age estimation works such as [19,21,22]. Han et al. [22] used selected BIF features in order to predict the age, gender and ethnicity attributes. Some researchers casted the age estimation problem into a label distribution learning (LDL) problem (e.g., [39,40]).

Due to their significant performance improvement in many image classification tasks, deep learning methods (e.g., [23,34]) have been proposed for age estimation in recent times. Deep learning methods claim to have the best performances in demographic attributes estimation (ethnicity, gender and age). It is known that deep learning paradigms can provide impressive results on a single database. However, when a new unseen database is used with the pre-trained deep net, the performance of age estimation can drop significantly. A retraining or a fine-tuning process can solve this problem. However, these tasks face many problems related to affordability and flexibility. Indeed, DCNNs are complicated models that require huge amount of labeled data and powerful computational facilities. More importantly, DCNNs are with many hyper-parameters, and the learning performance depends on careful tuning of them.

In this paper, we provide a comparative study of age estimation based on hand-crafted and deep features. More precisely, we focus on transfer learning. With transfer learning, a pre-trained deep CNN can be used as an image feature extractor. We will show that by adopting such transfer learning, the full power of a pre-trained DCNN can be exploited in order to process new unseen datasets without any retraining or fine-tuning. To this end, we deploy a shallow regressor (partial least square regressor) that is built and trained on top of the extracted deep features. Training this regressor is much more efficient than retraining or fine-tuning the whole deep net using a sheer number of images. Experiments will show that this scheme can efficiently adapt pre-trained networks to new unseen data. Besides, this scheme provides more accurate results than those obtained by the end-to-end deep net solution. These experiments also show that the transfer learning is independent of the type of ages (biological or apparent) used in the training phase of the original DCNN. The remainder of the paper is organized as follows: face alignment is briefly introduced in Sect. 2. The experimental setup is described in Sect. 3, and the evaluation of the results are given in Sect. 4. In Sect. 5 we give the conclusion.

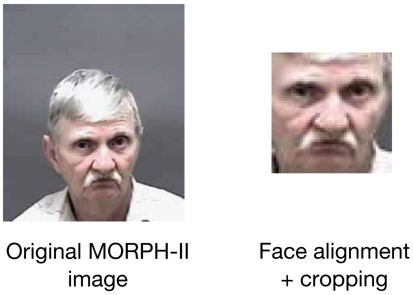

**Fig. 1** 2D face alignment and cropping associated with one original image in the MORPH database

## 2 2D face alignment

For image-based age estimation, face alignment constitutes one of the most important pre-processing stages. In our work, the eyes of each face are extracted using the ensemble of regression trees (ERT) method [25] which is a robust and fast algorithm for facial points localization. Once the 2D positions of the two eyes are estimated, the in-plane rotation can be compensated for. Finally, the rotated face image is rescaled such that the inter-ocular distance become equal to a fixed value $L$. After carrying out the rotation and rescaling, the face should be cropped. To this purpose, a bounding rectangle is centered on the new eyes location and then stretched in all four directions. Finally, in our study, the final face image has a size of $50 \times 50$ pixels for the MORPH dataset and $200 \times 200$ for the PAL dataset. Figure 1 shows the 2D alignment associated with one original face image in MORPH dataset.

## 3 Transfer learning and experimental setup

### 3.1 Face features

In our work, we use three hand-crafted features: local binary patterns (LBP) [1,3], histogram of oriented gradients (HOG) [8], and binarized statistical image features (BSIF) [24].

The adopted transfer learning is illustrated in Fig. 4. Obviously, transfer learning only concerns the deep CNNs. For transfer learning, we use deep features provided by five pre-trained Deep CNN: VGG-face [33], ImageNet [5], ImageNet VGG-verydeep-16 [41], DEX-IMDB-WIKI and DEX-ChaLearn-ICCV2015 [37,38]. We can note that the last two nets were trained on face images for the age estimation problem. The DEX-IMDB-WIKI network was learned using real ages using the aligned and cropped faces of the IMDB-WIKI dataset, while the DEX-ChaLearn is a fine-tuned version of the previous model, trained on apparent age using the challenge images. All deep nets are used as feature extractors. The LBP, HOG, and BSIF descriptors have 256,

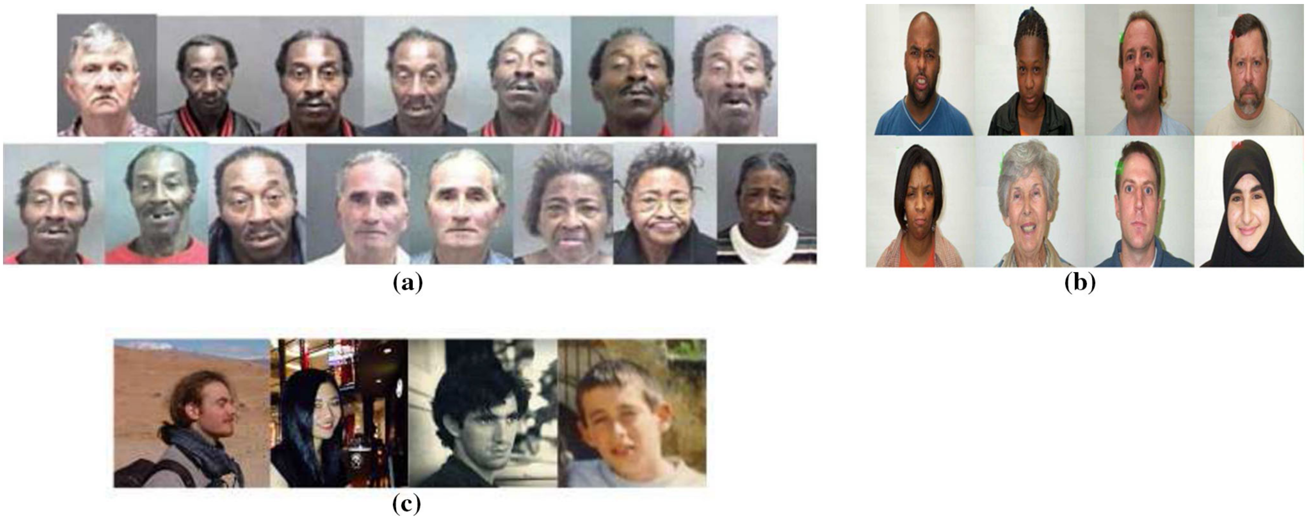

**Fig. 2** **a** Sample images from the MORPH database. **b** Sample images from the PAL database. **c** Sample images from the Chalearn2016 database

832/1872/4212,[1] and 256, respectively. All deep features are given by 4096 elements.

### 3.2 Datasets

In our study, three public datasets are used.

*MORPH (Album 2)* The MORPH (Album 2), or simply MORPH, database from the University of North Carolina Wilmington [35] has $\sim 55,000$ images of 13,618 persons (11,459 male and 2159 female) in the age range of 16–77 years. The mean number of images per person is 4. The MORPH database can be grouped into three main ethnicities: African (42,589 images), European (10,559 images) and other ethnicities (1986 images). Some images are shown in Fig. 2a. In our work, we use a fivefold cross-validation evaluation, and the folds are selected in such a way to prevent algorithms from learning the identity of the persons in the training set by making sure that all images of individual subjects are only in one fold at a time.

*PAL* The Productive Aging Lab Face (PAL) database from the University of Texas at Dallas [30] includes 1046 frontal face images from different subjects (430 males and 616 females) in the age range of 18–93 years. This database can be grouped into three main ethnicities: African-American subjects (208 images), Caucasian subjects (732 images) and other subjects (106 images). This database contains faces having different expressions. Some images are shown in Fig. 2b. For the evaluation of the approach, we use a fivefold cross-validation. In the experiments, we considered three scenarios: (i) original images, (ii) aligned images with loose crop (face plus some background), and (iii) aligned/cropped images. These scenarios are illustrated in Fig. 3. The corresponding sizes are 230 × 350 pixels, 200 × 200 pixels, and 200 × 200 pixels, respectively.

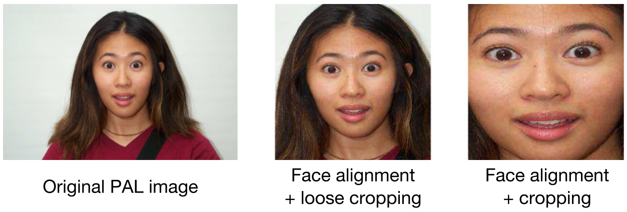

**Fig. 3** Three types of PAL images. The left one is the original face image. The middle and right images correspond to the aligned and cropped face. The middle image corresponds to a loose face cropping and the left one to a tight face cropping

*Chalearn2016* This database contains 8000 images labeled with the apparent age. Figure 2c shows some samples of the dataset. Each image has been labeled by multiple individuals, using a collaborative Facebook implementation and Amazon Mechanical Turk. The votes variance is used as a measure of the error for the predictions [10]. The mean is used as the ground-truth age. In our experiments, we use about 5341 images from the Chalearn2016 dataset. These images have available apparent ages. We note that the dataset used for training the DEX-ChaLearn network is different of the Chalearn2016 dataset.

### 3.3 Regression module

Since regression is used in the training and testing stage, deploying an efficient regression tool will be useful for both stages. Thus, we are interested in efficient shallow regressors. Age estimation can be considered as a univariate regression problem where the responses are given by the ages and the

[1] for original and aligned MORPH images and PAL images, respectively.

predictive variables are given by the image descriptors. In our work, we use the partial least square (PLS) regressor [36]. This is a statistical method that discovers relations between groups of observed variables $X$ and $Y$ via the use of latent variables. $X$ represents the observations and $Y$ the associated responses.

PLS is a powerful statistical method which can simultaneously perform dimensionality reduction and classification/regression with some similarities to PCA. While PCA reduces the dimensionality of a single data set by iteratively projecting the data onto components of maximum variance, PLS reduces the dimensionality of a data set by projecting the data onto components of maximum covariance with a second data set (responses $Y$)

For univariate problems, the PLS problem is referred to PLS1 problem. There are many algorithms for solving the PLS1 problem. More details can be found in [29].

In our work, we use the SIMPLS1 algorithm [9] which is generally faster than the original nonlinear iterative partial least squares (NIPALS) algorithm because it does not deflate the $X$ matrix. Instead, it deflates the matrix $X^T Y$ , which is usually much smaller in dimension than $X$.

SIMPLS1 calculates the PLS factors directly as linear combinations of the original variables. The PLS factors are determined such as to maximize a covariance criterion, while obeying certain orthogonality and normalization restrictions. The construction of deflated data matrices as in the NIPALS-PLS algorithm is avoided.

The MATLAB code that describes the steps of the SIMPLS1 algorithm are given in Algorithm 1.

**Algorithm 1** SIMPLS1

**Input:**
Centred $N$ data samples (block of predictors): $X \in \mathbb{R}^{N\times D}$;
Centred responses $y \in \mathbb{R}^N$ (a vector);
Number of latent features (PLS components): $M$
**Output:**
Regression model $\beta$;

```
function [β, R, T, P, Q] = SIMPLS1(X, y, M)

V = zeros(D,M);

s = X' * y; // X' denotes the transpose of X

for m = 1:M

r = s ;
t = X * r;
tt = norm(t);
t = t/tt;
r = r/tt;
p = X'*t;
q = y'*t; //q is a scalar
u = y*q;
v = p;

if m > 1

v = v - V * (V'*p);
u = u - T * (T'*u);
end

v = v/norm(v);
s = s - v*(v'*s);
R(:,m) = r;
T(:,m) = t;
P(:,m) = p;
Q(:,m) = q; // Q is a vector
U(:,m) = u;
V(:,m) = v;


end
β = R * Q'; // Regression parameters
```

### 3.4 Evaluation protocol

Figure 4 depicts the training and testing processes used for evaluating the performance obtained with the eight face features. The strategy is the same whether the features are hand-crafted or provided by the pre-trained DCNNs. We adopt fivefold cross-validation scheme that allows to test every test image in the studied database. For deep features, the training phase concerns the regressor only. As for evaluating the performance of a given feature, we use two measures that are very common in the literature for evaluating the performance of automatic age estimators.

The first measure is the mean absolute error (MAE) (expressed in years). It is computed as the average of absolute error between the predicted ages and the ground-truth ones. The MAE is given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \left|a_i - \hat{a}_i\right| \tag{1}$$

where $N, a_i, \hat{a}_i$ are the total number of images, the predicted age and the ground-truth age, respectively. The second measure is given by the cumulative score (CS). This score is equal to the percentage of tested images for which the age error is less than a given threshold. The CS is given by:

$$CS(l) = \frac{N_{e\leq l}}{N}\% \tag{2}$$

where $l$, $N$ and $N_{e\leq l}$ are the threshold (years), the total number of tested images, and the number of tested images, respectively, in which the age estimation error (in absolute value) is less than or equal to the threshold.

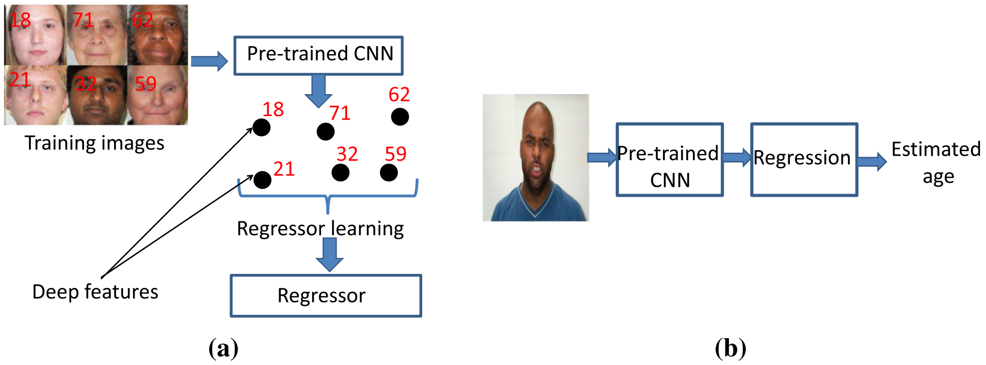

**Fig. 4** Transfer learning. **a** In the training phase, the training images are fed to a given pre-trained CNN in order to obtain the deep features. The deep features are the used for learning an age regressor. **b** For a test face, the age is predicted by using the learned regressor using the corresponding extracted deep features

**Table 1** Mean absolute error (years) obtained with different face features on MORPH database

| Face features | Images | Aligned + cropped |
|---|---|---|
| LBP | 7.20 | 6.53 |
| HOG | 6.26 | 4.84 |
| BSIF | 7.34 | 6.69 |
| VGG-Face | 4.72 | 4.79 |
| IMAGENET-VGG-F | 5.11 | 5.04 |
| IMAGENET-VERY-DEEP-16 | 5.53 | 5.47 |
| DEX-ChaLearn | **3.67** | 4.77 |
| DEX-IMDB-WIKI | 3.77 | **4.76** |

The numbers in bold represent the best results

**Table 2** MAE (years) obtained with different state-of-the-art approaches on MORPH database

| Publication | Approach | MAE |
|---|---|---|
| Fernandez et al. [11] | HOG + SVR | 4.8 |
| Guo and Mu [17] | BIF[a] + KPLS[b] | 4.2 |
| Chang et al. [4] | BIF[a] | 6.1 |
| Geng et al. [13] | BIF[a] | 4.8 |
| Guo and Mu [18] | BIF[a] | 4.0 |
| Huerta et al. [23] | CNN[c] | 3.9 |
| Han et al. [22] | DIF[d] | **3.6** |
| Wang et al. [42] | DLA | 4.7 |
| Proposed scheme | Transfer learning | 3.67 |

The number in bold represents the best results
[a] Biologically inspired features
[b] Kernel partial least squares
[c] Convolutional neural networks
[d] Demographic informative features

## 4 Experimental results

### 4.1 Single database

*MORPH* Table 1 depicts the MAE obtained on the MORPH database when the eight face features are used. In this table, we considered two scenarios: original images and the aligned/cropped images. As can be seen, performing face alignment and cropping has improved the performances of the hand-crafted features. The explanation is very intuitive. Hand-crafted features need to focus on the face region only. On the other hand, for the last two deep features, the use of the original images provided better performance. This can be explained by the fact that these two pre-trained DCNNs were trained on face images having significant background. For both scenarios (original images or the aligned and cropped images), the deep features provided by DEX-IMDB-WIKI, and DEX-ChaLearn nets provided the best performances. Moreover, we can observe that among deep features the best performances were obtained with nets that were trained on face image datasets, i.e. VGG-Face, DEX-IMDB-WIKI and DEX-ChaLearn.

The performance of some state-of-the-art approaches is summarized in Table 2. As it can be seen, the results obtained with transfer learning are comparable to the performance obtained by the work of Han et al. [22]. The latter uses coarse-to-fine and hierarchical age estimation adopting binary decision trees for classifying non-overlapping age groups and within-group age regressors. In our work, we only use one single regressor. Figure 5 illustrates the cumulative score associated with the eight face features. As can be seen, the top cumulative scores are similar.

Figure 6 shows the discrepancy between predicted and ground-truth ages in the MORPH database when the face features are provided by DEX-ChaLearn features. As it can be seen, at middle ages the age errors have some symmetry

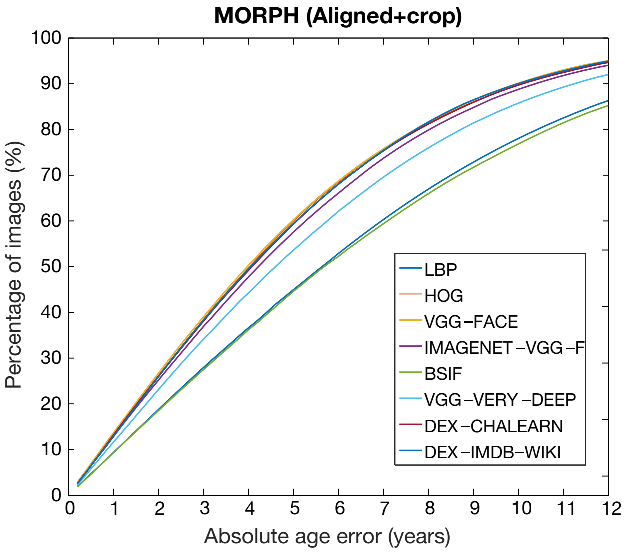

**Fig. 5** Cumulative scores obtained with eight face features for MORPH database (aligned and cropped images)

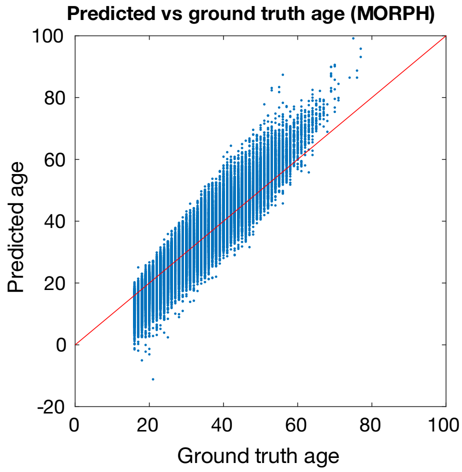

**Fig. 6** Predicted age versus ground-truth age for MORPH face images using DEX-ChaLearn features

with the respect to the corresponding ground-truth age. For the boundaries of the age range (i.e., young and old), the error symmetry was not observed anymore. Figure 7 shows the MAE for each age in the MORPH database when the face features are provided by DEX-ChaLearn features.

*PAL* Table 3 illustrates the MAE obtained on the PAL database using the eight face features. In this table, three scenarios were studied: (i) original images, (ii) aligned images with loose crop (face plus some background), and (iii) aligned/cropped images. As can be seen, face alignment and cropping helped hand-crafted features getting good performances. The performances obtained by the last two deep features were the best for all three scenarios (three types of cropping). In general, for the transfer learning adopting the five DCNNS, the best results were obtained when loose cropping is used.

The performance of some state-of-the-art methods are summarized in Table 4. As can be seen, by adopting the proposed transfer learning, we got a significant improvement in performance. The best state-of-the-art MAE was 5.4 years, whereas the best MAE obtained by our adopted scheme was 3.79 years. Figure 8 illustrates the cumulative score associated with the eight face features. Figure 9 illustrates some examples of good and poor age estimation in PAL database. The left column shows poor age estimation and the right column shows good age estimation. In Fig. 9, a poor estimation is declared for a test face if the absolute age error is equal to or greater than 1 year.

By comparing the obtained MAE using transfer learning of DEX-ChaLearn and DEX-IMDB-WIKI features in Table 1 (MORPH) as well as in Table 3 (PAL), we can observe that the transfer learning performance is almost the same for the two nets. Since the DEX-ChaLearn net was trained on apparent ages and the DEX-IMDB-WIKI net was trained on biological ages, we can conclude that the adopted transfer learning is independent of the age type (biological or apparent) used in the training of the original DCNNs.

**Fig. 7** MAE for each age in MORPH using the DEX-ChaLearn features

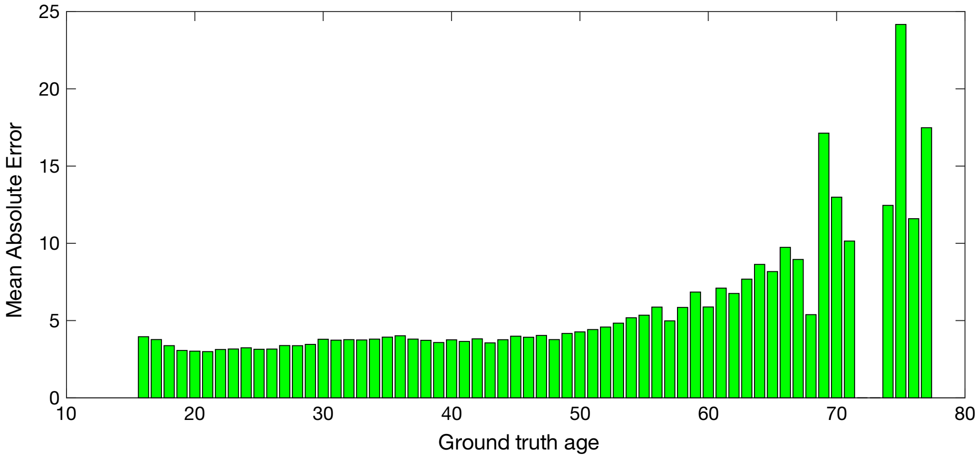

**Table 3** Mean absolute error (years) obtained with different face features on PAL database

| Face features | Original images | Aligned + Loose crop | Aligned + crop |
|---|---|---|---|
| LBP | 11.40 | 11.16 | 10.99 |
| HOG | 8.68 | 7.61 | 7.00 |
| BSIF | 10.71 | 11.26 | 10.09 |
| VGG-Face | 5.91 | 5.13 | 5.23 |
| IMAGENET-VGG-F | 6.89 | 6.81 | 7.14 |
| IMAGENET-VERY-DEEP-16 | 8.04 | 8.64 | 8.41 |
| DEX-ChaLearn | **3.97** | **3.79** | 5.12 |
| DEX-IMDB-WIKI | 4.04 | **3.79** | **4.90** |

The numbers in bold represent the best results

**Table 4** Mean absolute error (years) obtained with different state-of-the-art approaches on PAL database

| Publication | Approach | MAE |
|---|---|---|
| Gunay and Nabiyev [16] | AAM + GABOR + LBP | 5.4 |
| Nguyen et al. [32] | MLBP + GABOR + SVR | 6.5 |
| Nguyen et al. [31] | MLBP + GABOR + PCA + SVR | 6.0 |
| Bekhouche et al. [2] | LBP + BSIF + SVR | 6.2 |
| Choi et al. [6] | GHPF[a] + SVR | 8.4 |
| Luu et al. [28] | CAM[b] + SVR | 6.0 |
| Proposed scheme | Transfer learning | **3.79** |

The number in bold represents the best results

[a] Gaussian high pass filter

[b] Contourlet appearance model

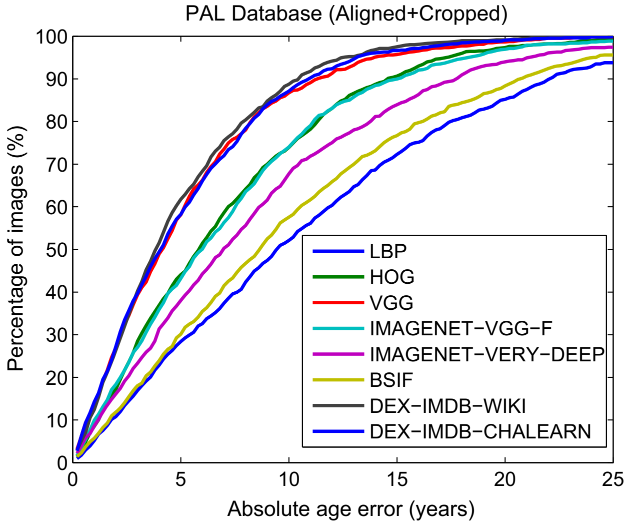

**Fig. 8** Cumulative scores obtained with eight face features for PAL database (aligned and cropped images)

*End-to-end use versus transfer learning* Since DEX-ChaLearn and DEX-IMDB-WIKI are deep CNNs that were trained to predict ages of input face images, we can compare their performance for age estimation with that obtained with transfer learning as described in the previous section. Table 5 shows a comparison between the MAE of the end-to-end DCNNs and that obtained by transfer learning. This table corresponds to the MORPH database with two different types of images: original and cropped. For each CNN, the upper row presents the MAE obtained by the end-to-end use. The lower row depicts the MAE obtained by transfer learning (use of deep features). As can be seen, by adopting transfer learning the obtained MAE was better than that of the end-to-end use of the DCNNs.

Table 6 shows a comparison between the MAE of the end-to-end DCNNs and that obtained by the use of transfer learning. This table corresponds to the PAL database with three different types of images. We can observe a similar behavior to that obtained with the MORPH database. This tends to confirm that by only retraining the regressor, we are able to transfer the power of the pre-trained CNN without having to retrain or fine-tune the whole network.

Table 7 depicts the MAE as a function of the number of latent components associated with the PLS regressor. We can observe that the use of 20 latent variables for almost all face features provided the best results.

## 4.2 Cross-database and cross-age-type evaluations

In order to have a quantitative cross-database evaluation, we conducted the following experiments. The goal is to study the performance of age estimation when training and test parts belong to different databases. In addition, we aim to study the performances when the training part depicts biological

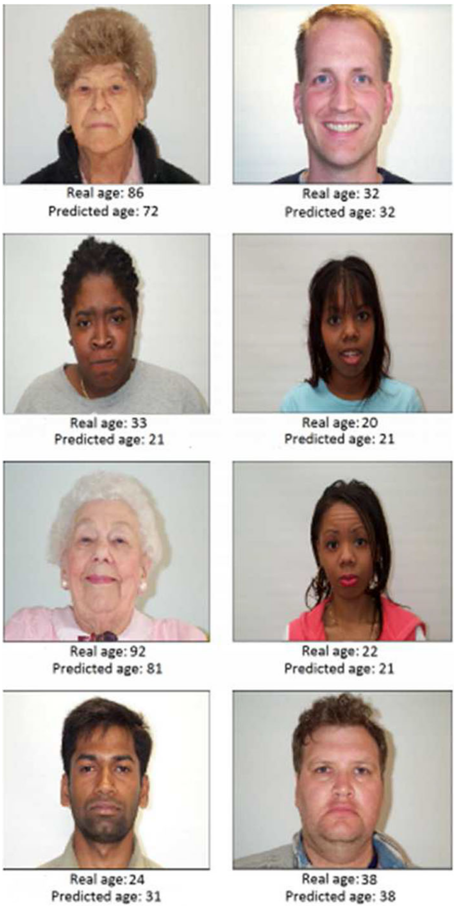

**Fig. 9** Examples of good and poor age estimation. The left column illustrates poor age estimation and the right column illustrates good age estimation

(apparent) age and the test part contains images with apparent (biological) ages.

To this end, we use the three databases MORPH, PAL, and Chalearn2016. Since the ground-truth ages for the MORPH and PAL databases correspond to biological ages and since the ground-truth ages for Chalearn2016 correspond to apparent ages, our conducted experiment will study the joint effect of database change and age type change. In this experiment, the protocol adopted is to use an entire database as the training set and the other as the testing set and vice versa. The results are summarized in Table 8. In this experiment, we use the deep features provided by the layer FC7 of DEX-chalearn. We use the PAL images (loose crop), original MORPH images and the cropped and aligned images for Chalearn2016. We can observe that when the PAL database is used as a training set (MORPH is used as a test set) the performance is better than the opposite case. The main reason relies in the size of the datasets used. Indeed, the PAL dataset has 1,046 images while the MORPH dataset has about 55,000 images. As a result, the trained model on the PAL dataset (when tested on MORPH) cannot be as accurate as the model trained on MORPH and tested on PAL dataset. Although both cases have adopted the cross-dataset scenario, the size of the training set can influence the final accuracy of the model.

Furthermore, the performance deteriorates when training/testing databases have different types of ages (biological and apparent ages). The deterioration was less severe when the training set used apparent ages than the opposite configuration.

As it can be seen, for Chalearn2016 face images, the MAE was ranging from 9 to 11.9 years. This large error is due to the significant difference in the databases as well as to the type difference of ground-truth ages. It is worthy to note that in the case of tested face images with apparent ages, the evaluation

**Table 5** Mean absolute error (years) obtained with two deep CNNs on MORPH database

| CNN | Scheme | Original | Aligned + crop |
|---|---|---|---|
| DEX-ChaLearn | End-to-end | 5.34 | 11.10 |
| | Transfer learning | **3.67** | **4.77** |
| DEX-IMDB-WIKI | End-to-end | 5.77 | 11.60 |
| | Transfer learning | **3.77** | **4.76** |

The numbers in bold represent the best results

For each CNN, the upper row illustrates the MAE obtained by applying the CNN as an end-to-end solution. The lower row depicts the MAE where the net is used to provide only the deep features

**Table 6** Mean absolute error (years) obtained with two deep CNNs on PAL database

| CNN | Scheme | Original | Aligned + loose crop | Aligned + crop |
|---|---|---|---|---|
| DEX-ChaLearn | End-to-end | 7.12 | 5.43 | 8.53 |
| | Transfer learning | **3.97** | **3.79** | **5.12** |
| DEX-IMDB-WIKI | End-to-end | 6.99 | 4.72 | 7.98 |
| | Transfer learning | **4.04** | **3.79** | **4.90** |

The numbers in bold represent the best results

**Table 7** MAE as a function of the latent variables used by the partial least square regressor

| Features\**nb. of PLS components** | 10 | 20 | 30 | 40 |
|---|---|---|---|---|
| LBP | 11.10 | **10.84** | 11.16 | 11.16 |
| HOG | **7.02** | 7.41 | 7.61 | 7.75 |
| BSIF | 11.33 | **10.91** | 11.26 | 11.34 |
| VGG-Face | **5.04** | 5.07 | 5.13 | 5.14 |
| IMAGENET-VGG-F | 6.65 | **6.45** | 6.81 | 7.10 |
| IMAGENET-VERY-DEEP-16 | 8.76 | **8.54** | 8.63 | 8.74 |
| DEX-ChaLearn | 3.79 | **3.74** | 3.79 | 3.87 |
| DEX-IMDB-WIKI | 3.84 | **3.79** | 3.79 | 3.94 |

The numbers in bold represent the best results
The results correspond to PAL database

**Table 8** The performance (MAE in years) of the cross-database evaluation in the case of age estimation

| Training | Testing | | |
|---|---|---|---|
| | Chalearn2016 | PAL | MORPH |
| Chalearn2016 | – | 9.04 | 9.18 |
| PAL | 11.91 | – | 6.75 |
| MORPH | 9.47 | 7.70 | – |

**Table 9** CPU time (ms) associated with the processing of one test image in three different datasets

| Dataset | Deep feature extraction (DExChaLEarn) | **PLS regression** | **Total** |
|---|---|---|---|
| Chalearn2016 | 47.7 | 0.04 | 47.7 |
| PAL | 45.7 | 0.04 | 45.7 |
| MORPH | 56.2 | 0.04 | 56.2 |

**Table 10** CPU time (s) associated with the training of PLS regressors for three different datasets

| Dataset | PLS training |
|---|---|
| Chalearn2016 (5341 images) | 0.663 |
| PAL (1046 images) | 0.213 |
| MORPH (55,134 images) | 5.623 |

can also use another measure which takes into account the standard deviation of the ground-truth age. This is a given by:

$$\epsilon = 1 - \frac{1}{N}\sum_{i=1}^{N} e^{-\frac{(a_i-\hat{a}_i)^2}{2\sigma_i^2}} \tag{3}$$

where $a_i$ denotes the estimated age and $\hat{a}_i$ the ground truth apparent age. $\sigma_i$ is the standard deviation associated with the apparent age $a_i$. The above formula is not used in Table 8 since the cross-database evaluation involves apparent and biological ages.

## 4.3 Computational complexity

We emphasize that the CPU time of proposed age estimation methods is not always reported in the literature. Indeed, most researchers in this field were focusing on improving the accuracy of their methods and systems. They did not report the computational complexity of their proposed methods. Nevertheless, for completeness of results, this section will depict the execution time needed for age estimation. To this end, we fix the face features to the deep features provided by DEX-ChaLearn networks.

The CPU times in ms are illustrated in Table 9. The first column depicts the CPU time associated with deep feature extraction. The second column depicts the CPU time associated with the linear PLS regression. The specifications of the used hardware and software are as follows.

Processor: Intel Core i7-5960X CPU @ 3.00GHz x 8
Memory: 64 GB
Graphics Card: NVIDIA Corporation GM107GL [Quadro K2200]
Software: Caffe and Matcaffe (downloaded from GitHub) are using CUDA-7.5, OpenCV3.1.0 and MATLAB R2016a.

We also measured the CPU time needed for the PLS regressor training over the entire datasets. Table 10 summarizes these times in seconds when the number of latent variables is set to 30.

# 5 Conclusion

This paper has addressed age estimation in facial images using transfer learning. We have compared several face features for the task of age estimation. In the study, we have considered three shallow image features as well as five deep features provided by pre-trained DCNNs. The comparison shown in the paper yields several conclusions. First, by adopting transfer learning of DCNNs, efficient and accurate age estimation can be obtained on new datasets on the premise that the age regressor is retrained. The adopted transfer learning is by far more efficient than retraining the whole deep CNN on the new set of images. Second, the use of deep features gave better results than those obtained with handcrafted features. Third, the accuracy obtained by transfer learning can be highly correlated to the deep net context (object type in the training data, the targeted problem, etc.). Fourth, transfer learning is still successful regardless of the age type used for training the DCNN, whereas a full cross-age type may lead to a performance drop. Fifth, in some cases, the performance can be improved by using loose face cropping.

**Acknowledgements** The authors gratefully acknowledge the NVIDIA Corporation for the donation of the Titan X used for this research.

## References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
2. Bekhouche, S., Ouafi, A., Taleb-Ahmed, A., Hadid, A., Benlamoudi, A.: Facial age estimation using BSIF and LBP. In: International Conference on Electrical Engineering (2014)
3. Bereta, M., Karczmarek, P., Pedrycz, W., Reformat, M.: Local descriptors in application to the aging problem in face recognition. Pattern Recognit. **46**, 2634–2646 (2013)
4. Chang, K.Y., Chen, C.S., Hung, Y.P.: Ordinal hyperplanes ranker with cost sensitivities for age estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 585–592 (2011)
5. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: British Machine Vision Conference (2014)
6. Choi, S.E., Lee, Y.J., Lee, S.J., Park, K.R., Kim, J.: A comparative study of local feature extraction for age estimation. In: International Conference on Control Automation Robotics Vision, pp. 1280–1284 (2010)
7. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2005)
9. de Jong, S.: Simpls: an alternative approach to partial least squares regression. Chemom. Intell. Lab. Syst. **18**(3), 251–263 (1993)
10. Escalera, S., Torres, M., Martinez, B., Baró, X., Escalante, H.J., et al.: Chalearn looking at people and faces of the world: face analysis workshop and challenge 2016. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition Workshops (2016)
11. Fernandez, C., Huerta, I., Prati, A.: A comparative evaluation of regression learning algorithms for facial age estimation. In: FFER in Conjunction with IEEE Int. Conf. on Pattern Recognition (2014)
12. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. IEEE Trans. Multimed. **10**(4), 578–584 (2008)
13. Geng, X., Yin, C., Zhou, Z.H.: Facial age estimation by learning from label distributions. IEEE Trans. Pattern Anal. Mach. Intell. **35**(10), 2401–2412 (2013)
14. Geng, X., Zhou, Z.H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2234–2240 (2007)
15. Gunay, A., Nabiyev, V.V.: Automatic detection of anthropometric features from facial images. In: 2007 IEEE 15th Signal Processing and Communications Applications, pp. 1–4 (2007)
16. Günay, A., Nabiyev, V.V.: Age estimation based on hybrid features of facial images. In: Information Sciences and Systems 2015: 30th International Symposium on Computer and Information Sciences (ISCIS 2015), pp. 295–304. Springer, Cham (2016)
17. Guo, G., Mu, G.: Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 657–664 (2011)
18. Guo, G., Mu, G.: Joint estimation of age, gender and ethnicity: CCA vs. PLS. In: IEEE International Conference and Workshop on Automatic Face and Gesture Recognition, pp. 1–6 (2013)
19. Guo, G., Mu, G.: A framework for joint estimation of age, gender and ethnicity on a large database. Image Vis. Comput. **32**(10), 761–770 (2014)
20. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: Computer Vision Pattern Recognition (2009)
21. Han, H., Jain, A.K.: Age, gender and race estimation from unconstrained face images. Technical Report MSU-CSE-14-5, Department of Computer Science, Michigan State University, East Lansing, Michigan (2014)
22. Han, H., Otto, C., Liu, X., Jain, A.K.: Demographic estimation from face images: human vs. machine performance. IEEE Trans. Pattern Anal. Mach. Intell. **37**(6), 1148–1161 (2015)
23. Huerta, I., Fernandez, C., Segura, C., Hernando, J., Prati, A.: A deep analysis on age estimation. Pattern Recognit. Lett. **68**(2), 239–249 (2015)
24. Kannala, J., Rahtu, E.: BSIF: binarized statistical image features. In: 2012 21st International Conference on Pattern Recognition (ICPR), pp. 1363–1366 (2012)
25. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
26. Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 762–767 (1994)
27. Levi, G., Hassncer, T.: Age and gender classification using convolutional neural networks. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 34–42 (2015)
28. Luu, K., Seshadri, K., Savvides, M., Bui, T.D., Suen, C.Y.: Contourlet appearance model for facial age estimation. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–8 (2011)
29. Martin, A.: A comparison of nine pls1 algorithms. J. Chemom. **23**(10), 518–529 (2008)
30. Minear, M., Park, D.C.: A lifespan database of adult facial stimuli. Behav. Res. Methods Instrum. Comput. **36**(4), 630–633 (2004)
31. Nguyen, D.T., Cho, S.R., Pham, T.D., Park, K.R.: Human age estimation method robust to camera sensor and/or face movement. Sensors **15**(9), 21898–21930 (2015)
32. Nguyen, D.T., Cho, S.R., Shin, K.Y., Bang, J.W., Park, K.R.: Comparative study of human age estimation with or without preclassification of gender and facial expression. Sci. World J. **2014**, 15 (2014)
33. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference, vol. 1, p. 6 (2015)

34. Ranjan, R., Zhou, S., Chen, J.C., Kumar, A., Alavi, A., Patel, V.M., Chellappa, R.: Unconstrained age estimation with deep convolutional neural networks. In: 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 351–359 (2015)
35. Ricanek, K., Tesafaye, T.: MORPH: a longitudinal image database of normal adult age-progression. In: 7th International Conference on Automatic Face and Gesture Recognition, 2006, pp. 341–345. FGR 2006 (2006)
36. Rosipal, R., Kramer, N.: Overview and recent advances in partial least squares. In: Subspace, Latent Structure and Feature Selection Techniques, pp. 34–51. Springer (2006)
37. Rothe, R., Timofte, R., Gool, L.V.: DEX: deep expectation of apparent age from a single image. In: IEEE International Conference on Computer Vision Workshops (ICCVW) (2015)
38. Rothe, R., Timofte, R., Gool, L.V.: Deep expectation of real and apparent age from a single image without facial landmarks. Int. J. Comput. Vis. (IJCV) **126**, 144–157 (2016)
39. Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., Yuille, A.: Deep regression forests for age estimation (2017). https://arxiv.org/pdf/1712.07195.pdf
40. Shen, W., Zhao, K., Guo, Y., Yuille, A.: Label distribution learning forests. In: NIPS (2017)
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:1409.1556
42. Wang, X., Guo, R., Kambhamettu, C.: Deeply-learned feature for age estimation. In: IEEE Workshop on Applications of Computer Vision (2015)
43. Zhang, Y., Li, L., Li, C., Cheng, C.: Quantifying facial age by posterior of age comparisons. In: BMVC (2017)

**F. Dornaika** received the M.S. degree in signal, image and speech processing from Grenoble Institute of Technology, France, in 1992, and the Ph.D. degree in computer science from Grenoble Institute of Technology, France in 1995. He is currently a Research Professor at IKERBASQUE (Basque Foundation for Science) and the University of the Basque Country. Prior to joining IKERBASQUE, he held numerous research positions in Europe, China, and Canada. He has published more than 250 papers in the field of computer vision and pattern recognition.

**I. Arganda-Carreras** is an Ikerbasque Research Fellow at the Computer Science and Artificial Intelligence department of the Basque Country University, Spain. He earned his Ph. D. in Computer Science and Telecommunications at the Universidad Autonoma de Madrid, Spain, in 2009. He was a postdoctoral fellow at the Massachusetts Institute of Technology, USA from 2009 to 2013, and at INRA, France, from 2013 to 2015.

**C. Belver** obtained his bachelor’s degree in Computer Science from the Basque Country University in 2015. He earned his Master degree in Computer Engineering and Intelligent Systems at the Basque Country University, Spain, in 2016.