Signal Processing: Image Communication 78 (2019) 368–375

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

DeepAge: Deep Learning of face-based age estimation ☆

Omry Sendik, Yosi Keller *

Faculty of Engineering, Bar Ilan University, Israel

A B S T R A C T

The estimation of a person's age based on a face image is a common biometric task conducted effortlessly by human observers. We present a dual Convolutional Neural Network (CNN) and Support Vector Regression (SVR) approach for face-based age estimation. A CNN is trained for representation learning, followed by Metric Learning, after which SVR is applied to the learned features. This allows to overcome the lack of large datasets with age annotations, by initially training the CNN for face recognition. The proposed scheme was applied to the MORPH-II and FG-Net datasets and compares favorably with contemporary state-of-the-art approaches. In particular, we show that domain adaptation which is essential for analyzing small-scale datasets, such as the FG-Net, can be achieved by retraining the SVR layer, rather than the CNN.

1. Introduction

The estimation of biometric traits based on face images is a common task for human observers that can often estimate the identity, age, gender, ethnicity, and kin relations of human subjects based on their face image. The derivation of such computational approaches has attracted significant research efforts, by studying face recognition [1,2], gender classification [3] and kinship verification [4,5], to name a few. In this work we study face based age estimation, where given a face image we aim to estimate the subject's age a, as depicted in Fig. 1.

Face-based age estimation is commonly formulated as either a classification problem, where an age corresponding to the face $\mathbf{x}$ is classified as either one of $\mathbf{a} = \{a_i\}_1^c$ discrete values [6–9], or as a regression problem, such that $a \in \mathbb{R}^+$ [6,10–14]. The common approach to face-based biometric analysis is to align the face image to a canonical spatial frame [15], and encode the face image using general purpose image descriptors, such as HOG and LBP [3], or face-specific descriptors [16]. This results in high dimensional representations that are used for recognition by Kernel SVM [9,17] or regression by Kernel PLS [7]. Convolutional Neural Networks (CNNs) allow to circumvent the need for handcrafted face descriptors, by optimizing task specific descriptors that are implicitly learnt through the CNN training, and are manifested by the output of the convolution layers of the CNN. Such approaches were also applied to face-based age estimation [11,13].

In this work we propose an *age regression* scheme, that learns a global face representation $\phi \in \mathbb{R}^d$ of the input face image via a CNN and employs a regression model using Support Vector Regression (SVR) [18], relating the learnt representation ϕ to the subject's age a. We show that Kernel-based SVR regression is improved by applying Metric Learning (ML) to the CNN-based representation ϕ, by learning a Mahalanobis distance

$$d^2_{\mathbf{W}}(\phi_i, \phi_j) = \left\| \mathbf{W}\phi_i - \mathbf{W}\phi_j \right\|_2^2, \tag{1}$$

such that $\hat{\phi} = \mathbf{W}\phi \in \mathbb{R}^{\hat{d}}$, $\hat{d} \ll d$ is a low dimensional face representation encoding the age variability, and face images related to similar ages will be closer together in the L_2 sense, while the L_2 distance between dissimilar face increases.

As the SVR is applied using the Radial Basis Function (RBF) kernel

$$K\left(\hat{\phi}_i, \hat{\phi}_j\right) = \exp\left(-\frac{1}{\sigma^2}\left\|\hat{\phi}_i - \hat{\phi}_j\right\|_2^2\right), \tag{2}$$

the improved separation of similar/dissimilar faces in the projected space $\hat{\phi}$, improves the SVR accuracy, as depicted in Fig. 2, and is further discussed in Section 3.4 .

In particular, we show that the proposed approach outperforms an end-to-end trained CNN, with the same architecture, that utilizes linear regression as its loss layer. This reveals a particular case where hybrid nonlinear-CNN schemes such as our (CNN and SVR), outperform end-to-end CNN-based regression due to the superior performance of nonlinear schemes that cannot be incorporated into the CNN backpropagation.

Thus, we present the following contributions:

First, we derive a fully data-driven approach for age estimation based on face images, consisting of a CNN-based representation learning trained for face recognition and age regression, followed by SVR. This allows to overcome the lack of large-scale age regression training sets, and utilize existing large-scale face recognition datasets such as the DeepFace dataset [19], and corresponding pretrained CNNs.

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.image.2019.08.003.

* Corresponding author.
E-mail addresses: omrysendik@gmail.com (O. Sendik), yosi.keller@gmail.com (Y. Keller).

https://doi.org/10.1016/j.image.2019.08.003
Received 9 April 2019; Received in revised form 21 June 2019; Accepted 4 August 2019
Available online 9 August 2019

(a) Age 32 (b) Age 11 (c) Age 88

Fig. 1. Face-based age estimation. Given an input face image, we aim to estimate the subject's age.

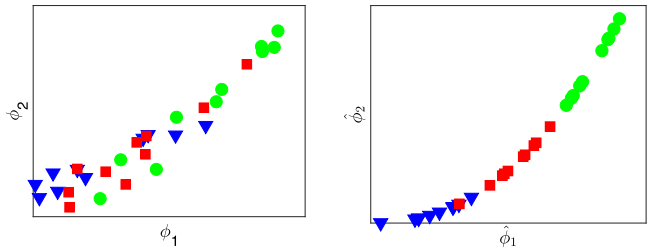

Fig. 2. The application of age-adaptive Metric Learning to nonlinear scalar regression. The different colors and shapes of the samples depict different ages. (a) Before the use of Metric Learning the different samples are partially mixed, due to multiple latent driving factors such as age, gender and ethnicity. Hence, applying a regression scheme might result in inaccurate age estimation. (b) The age-adaptive Metric Learning allows to better separate the samples in the features face with respect to their age.

Second, we show that the CNN-based representation can be further improved by applying regression-specific Metric Learning to derive a more discriminative low-dimensional face representation that allows to improve the SVR age estimate. Moreover, the Metric Learning and SVR can be retrained using small datasets compared to the CNN, and can thus be effectively utilized for domain adaptation that is essential in analyzing small-scale datasets such as the MORPH-II and FG-Net [20].

Last, The resulting scheme is shown to outperform state-of-the-art age estimation schemes, when applied to contemporary state-of-the-art face datasets. It also outperforms as end-to-end CNN-based scheme based on the same CNN architecture.

The rest of this paper is organized as follows: We present prior work on face-based age estimation in Section 2, and the proposed approach is introduced in Section 3. Experimental results and implementation issues are reported in Section 4, while conclusions are discussed in Section 5.

2. Related work

Face-based age estimation is a common trait of human vision despite varying facial aging characteristics, due to cultural-genetic and cranio-facial variations across ethnicities and genders. Age estimation is commonly formulated as either a classification problem, where a face is classified as being related to one of $\{a_1, a_2, \ldots, a_N\}$ ages, or as a regression problem aiming to estimate the age as a scalar $a \in \mathbb{R}^+$.

A classification based approach was proposed by Guo and Mu [21], based on a two-step procedure where gender and ethnicity are first classified, and the age is estimated separately for each gender and ethnicity group. The authors also derived a Kernel PLS based scheme [7] for age, race and gender estimation evaluated on the MORPH-II [22] dataset.

The OHRank scheme by Chang and Chen [8] utilized ordinal classification for age estimation, where each ordinal hyperplane separates all of the facial images into two groups according to the relative order, and the ages are inferred by aggregating a set of references from the ordinal hyperplanes. Ranking was also used by Zheng and Sun [9] who proposed the Ranking SVM scheme, where age is estimated indirectly, by first learning the age ranking relationships, and then estimating the age based on the ranking relationship and the ages in a reference set.

Choir and Lee [17] proposed a hierarchical age classification scheme that utilizes both global and local facial features. The classifier is based on SVM and SVR and uses overlapping age groups by considering the acceptance and rejection errors of each classifier. Similarly, Han and Otto [23] proposed a hierarchical approach for automatic age estimation, and provided an analysis of the influence of aging on individual facial features.

A different school of thought formulated age estimation as a scalar regression problem. Thus, Lanitis et al. [24] applied quadratic regression to map the Active Appearance Model (AAM) facial features to an age. They also derived a statistical model of facial appearance, utilized to derive a compact parametrization of face images. The Yamaha Gender and Age (YGA) database was studied by Fu et al. [25], who applied manifold learning to age estimation by modeling the low-dimensional manifold of face images using multiple linear regression functions. Discriminative subspace learning was applied by the same authors [26] to estimate aging patterns and apply multiple linear regressors with a quadratic cost function.

The Spatially Flexible Patch (SFP) descriptor was presented by Yan et al. [16] who encode both local appearance and position. The SFP relates to a particular age label by a Gaussian Mixture Model (GMM), and the age estimation is conducted by maximizing the sum of likelihoods related to the SFPs associated with a particular age. Similarly, Yan et al. [27] presented a patch-based regression framework for age and head pose estimation, where each image was encoded as a set of patches, whose distribution was modeled by a GMM, and the age and pose were estimated by kernel regression. Chen and Gong [12] introduced a cumulative attribute for learning a regression model, where sparse and imbalanced data are available to estimate age and crowd density. Another sparse regression model proposed by Demontis et al. [28] was trained using the FRGC (Face Recognition Grand Challenge) dataset [29], and was applied to the FG-Net dataset. A large dataset of 391 K human aging image was collected by Ni et al. [30] to train an age estimation system, using a multi-instance kernel-based regression scheme, and the resulting age estimator was applied to the FG-Net dataset.

As face image representations are high dimensional, low dimensional embedding can be applied. Thus, Guo et al. [31] applied the LARR (Locally Adjusted Robust Regression) scheme that learns a low-dimensional embedding of the aging manifold for age estimation, and also applied Marginal Fisher Analysis (MFA) and Locality Sensitive Discriminant Analysis (LSDA) [32]. Chao and Liu [10] utilized Metric Learning for dimensionality reduction, to encode face features computed by Active Appearance Models (AAM).

CNN-based schemes forgo the use of handcrafted image descriptors. Thus, Wang et al. [11] proposed a hierarchical unsupervised neural network architecture that first utilizes a CNN to learn low-level translation invariant features which are used as inputs to a set of Recurrent Neural Networks (RNNs). Manifold learning is applied to capture the underlying face aging structure by projecting the feature vector into a new low-dimensional more discriminative subspace. Marginal Fisher Analysis and Orthogonal Locality Preserving Projections were applied for dimensionality reduction, while SVM and SVR were used for age classification and regression, respectively. A CNN-based age estimation approach, similar to ours, was previously proposed by Wang et al. [33], where a CNN is trained using a Softmax loss from scratch with respect to age classes, and multiple layers are used for Representation Learning. Multiple manifold learning and regression schemes are then applied using the learnt features. In contrast, in the proposed scheme, we start by training a face recognition CNN based on classifying face images, thus allowing to utilize pre-trained face recognition CNNs, and show that Metric Learning is particularly related to Kernel-based schemes such as Kernel-SVR. This allows the proposed scheme to outperform both [11] and [33], when applied to the MORPH-II dataset.

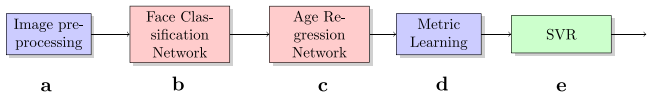

Fig. 3. The training pipeline of the proposed DeepAge face-based age estimation scheme. (a) The input images are preprocessed. (b) A face classification CNN (FCN) is trained. (c) the FCN is refined by replacing its loss function with L_2 regression and retraining an Age Regression Network (ARN). (d) Metric learning is applied to the FC layers of the ARN, and an age adaptive image representation $\hat{\phi}$ is computed. (e) Support Vector Regression (SVR) is applied to $\hat{\phi}$.

Metric Learning for face recognition was incorporated in an end-to-end CNN scheme by Schroff et al. [1]. They proposed to apply a Triplet Loss that minimizes the distance between an anchor face image and a positive image, both relating to the same identity, and maximizes the distance between the anchor and a negative of a different identity. State-of-the-art face recognition accuracy was achieved using a 128-bytes long descriptor. A triplet-based Metric Learning scheme for patch correspondence was proposed by Choy et al. [34], where an emphasis was put on hard mining of negative samples and a Hinge Loss was only applied to the negative part of the triplet loss.

A Metric Learning scheme similar to ours was applied by Yi et al. [35] to the detection and descriptors extraction of feature points in images using an end-to-end CNN-based approach. Hassner and Levi [13], and Yi et al. [14] reported significant accuracy increase by formulating the age estimation as a classification problem, and applying state-of-the-art CNNs.

The Chalearn dataset [36] was introduced as a large scale age estimation challenge, consisting of 7591 images with *apparent age* rather than real age. Such that the annotations were gathered using a Facebook app presented to Facebook users as a game, without a real baseline, as in the MORPH-II dataset, or any accuracy estimate.

Motivated by the recent improvements achieved through the use of CNNs, and the past performance of Metric Learning, we propose a method employing both, which is shown to outperform previous state-of-the-art approaches.

3. Deepage age estimation

The proposed DeepAge approach is an *age regression* scheme, that given a subject's face image $\mathbf{x}$ aims to estimate the subject's age a. Due to the limited size of face datasets with age annotations, it might be difficult to train a CNN from scratch. In contrast, there are multiple large scale face datasets used for face recognition, such as the DeepFace dataset [19] and corresponding pretrained CNNs, that we propose to utilize using the algorithmic framework depicted in Fig. 3.

We start by training a CNN to classify face images to one of C subjects, same as in face recognition tasks, and denote this CNN as Face Classification Network (FCN), as detailed in Section 3.2 and depicted in Fig. 3b. In the second step (Fig. 3c), we apply Transfer Learning by refining the FCN network by training it for an age regression task using a L_2 loss, and denote this network Age Regression Network (ARN). Although the ARN can be directly applied to age estimation, the gist of our approach is to show that the ARN's age estimation accuracy can be improved by using one of its Fully Connected (FC) layers $\phi_i \in \mathbb{R}^d$ as an age-adapted face descriptor. This representation is further refined by applying Metric Learning using a age-tagged face images, that explicitly maximizes the separation (in the L_2 sense) between the representations related to different ages (Section 3.3 and Fig. 3d). The resulting representation $\hat{\phi} \in \mathbb{R}^{\hat{d}}$, such that $\hat{d} \ll d$, is used by a Kernel-based Support Vector Regression (SVR) to estimate a_i in Fig. 3e, and is shown to outperform the age estimation accuracy of the end-to-end trained ARN.

The proposed scheme is summarized in Algorithm 1, corresponding to the steps in Fig. 3.

Algorithm 1 The proposed age estimation scheme

1: Alignment of the input face image.
2: Training a Face Classification Network (FCN) using a Softmax loss.
3: Refining the FCN by retraining it using a L_2 loss to derive the Age Regression Network (ARN).
4: Training and applying Metric Learning to either the $FC6$ or $FC7$ layer of the ARN to derive an age-adaptive descriptor $\phi \in \mathbb{R}^d$.
5: Training and applying SVR to ϕ to compute the age estimate.

3.1. Preprocessing of the face images

The application of a CNN to an image requires pre-processing to account for appearance and geometrical variations (Fig. 3a). For that we applied the Steepest Descent Method (SDM) [15] to detect facial landmarks in the face images. We chose a subset of nine feature points detected in all face images and aligned them to a canonical face by applying an affine transform. The canonical face was computed by first centering all faces and scaling them to a unit height. The images were than normalized by subtracting the average color value per pixel to account for photometric and appearance variations.

3.2. Representation learning

In order to derive an age-adaptive face representation using the relatively small age-annotated MORPH-II set [22] consisting of $\sim 60k$ images, we propose to apply Transfer Learning, where the CNN is initially trained for face recognition and then refined to derive an age regression Network (ARN) that is used for representation learning. Face recognition networks are typically trained for face classification [1,19] using a set of face images, where one of the Fully Connected (FC) layers is used as a face descriptor.

We propose two *alternative* implementations of the Face Classification Network (FCN), where the subsequent steps are applied mutatis mutandis to either of the two FCNs. The training of the FCN is depicted in Fig. 3b and step #2 in Algorithm 1.

The first FCN is the *pretrained* VggFace [19] CNN, that is considered state-of-the-art in face recognition and was trained using 2.6M face images. The VggFace CNN cannot be used to compare against most previous works in face-based age estimation that were trained using only the MORPH-II set [22]. Hence, we trained the second FCN *from scratch* based on the AlexNet [37] CNN, using only a subset of the MORPH-II set. AlexNet was chosen as one of the simplest CNN architectures, to emphasize the applicability of the proposed scheme, and was trained using a Softmax loss applied to a 60%/20%/20% training/validation/testing split of the MORPH-II set. It is depicted in Fig. 4 as the CNN whose input is a face image and an identity label, using a Softmax loss.

In order to adapt the FCN to age estimation, we refine its FC layer using a L_2 regression loss with respect to the subjects' ages using the same 60%/20%/20% split used for the FCN training. The resulting Age Regression Network (ARN) is also depicted Fig. 4 utilizing a L_2 regression loss. An ARN can be applied directly to age estimation. However, the gist of our approach is to show that it is better, in terms of age estimation accuracy, to utilize the ARN as an age-adaptive face representation $\phi \in \mathbb{R}^d$, $d = 4096$ that is further refined by Metric Learning (in Section 3.3) to yield a lower-dimensional representation $\hat{\phi} \in \mathbb{R}^{\hat{d}}$ used by a Support Vector Regression (SVR) scheme.

3.3. Age adaptive metric learning

Given the CNN-based representation $\phi_i \in \mathbb{R}^d$ of an input image $\mathbf{x}_i$, SVR can be applied directly to ϕ_i, where $d = 4096$, implying that the SVR training requires a large training set and significant computational complexity. It is common to apply dimensionality reduction via PCA to

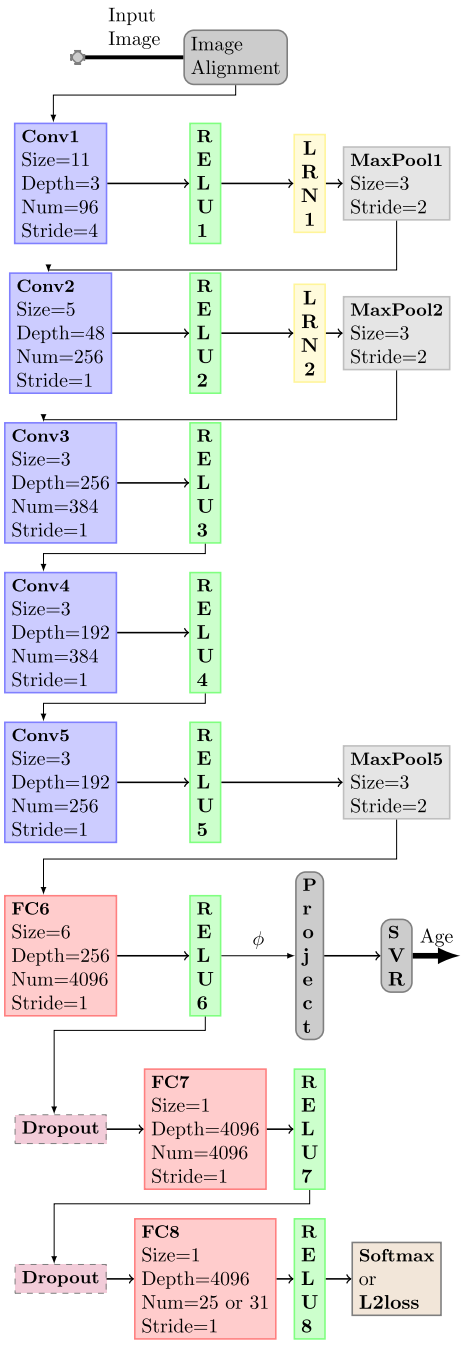

Fig. 4. An overview of the proposed scheme. The Face Classification Network (FCN) and Age Regression Network (ARN), are trained using a Softmax and L_2 losses, respectively. Either FC6 or FC7 can be used as input to the Metric Learning-bases projection layer and SVR, that are trained separately. The CNN is based on the AlexNet architecture [37].

reduce the SVR computational complexity and avoid over-fitting. For a given resulting PCA dimensionality $\hat{d} \ll d$ the PCA optimally retains the L_2 distances between samples.

We propose to apply supervised dimensionality reduction via Metric Learning as in Eq. (1) such that $\mathbf{W} \in \mathbb{R}^{\hat{d} \times d}$, $\hat{d} \ll d$, aiming to further improve the age estimation accuracy. The computational complexity of applying PCA and Metric Learning projections during the test phase is identical, as both are implemented by multiplying $\phi_i \in \mathbb{R}^d$ by the pre-computed projection matrix. The schemes differ in the offline training of the projection matrix.

The training is achieved by maximizing the classification margin via a hinge loss

$$(\mathbf{W}, \mathbf{b}) = \arg\min_{\mathbf{W},\mathbf{b}} \sum_{i,j} \max[r - y_{i,j}\left(b - \left(\phi_i - \phi_j\right)^T \mathbf{W}^T \mathbf{W}\left(\phi_i - \phi_j\right)\right), 0] \tag{3}$$

where b is the mean distance among all samples, r specifies the classification margin with respect to b, such that the classification margin between positive and negative samples is $2r$. The labels are given by

$$\begin{cases} |a_i - a_j| \le T & same \\ |a_i - a_j| > T & not\ same \end{cases} \tag{4}$$

such that a_i is the age of the i'th sample, and T is a predefined age difference threshold.

Eq. (3) is solved by a stochastic gradient descent (SGD), where at each iteration a pair of images is randomly drawn from the same/not same training sets, and the update is given by

$$\mathbf{W}_{t+1} = \begin{cases} \mathbf{W}_t & y_{ij}(b - \left(\phi_i - \phi_j\right)^T \mathbf{W}_t^T \mathbf{W}_t \left(\phi_i - \phi_j\right)) > r \\ \mathbf{W}_t - \gamma y_{ij} \mathbf{W}_t \left(\phi_i - \phi_j\right)\left(\phi_i - \phi_j\right)^T, & else \end{cases} \tag{5}$$

where $y_{ij} = 1$ and $y_{ij} = -1$ for positive and negative examples, respectively. b is initialized as the mid point between the average distances of subjects with similar and dissimilar classes, and γ is the predefined SGD learning rate.

3.4. Age estimation using support vector regression

Age estimation is a regression problem, where given a metric representation $\hat{\phi} \in \mathbb{R}^{\hat{d}}$ we aim to estimate a scalar function $a\left(\hat{\phi}\right) \in \mathbb{R}^+$. This is a supervised learning problem where we are given a training set of tuples $\left\{\hat{\phi}_i, a\left(\hat{\phi}_i\right)\right\}$. Kernel-based Support Vector Regression (SVR) [18] is the state-of-the-art approach for such problems, that was shown to outperform other regression schemes when applied to face-based age estimation [11,33]. Similar to SVM classification, the SVR estimate $a\left(\hat{\phi}\right)$ is given by [18]

$$f\left(\hat{\phi}\right) = \sum_{k=1}^{K} \alpha_k K\left(\hat{\phi}, \hat{\phi}_k\right) + b \tag{6}$$

where $\left\{\hat{\phi}_k, \alpha_k\right\}_1^K$ are the support vectors and corresponding weights. Kernel SVR is commonly applied using a Radial Basis Function (RBF) kernel

$$K\left(\hat{\phi}_i, \hat{\phi}_j\right) = \exp\left(-\frac{1}{\sigma^2}\left\|\hat{\phi}_i - \hat{\phi}_j\right\|_2^2\right). \tag{7}$$

The use of Metric Learning, as detailed in Section 3.3, improves the accuracy of Kernel-SVR by better separating, in the L_2 sense, between samples related to different ages. For instance, assume the ideal case where corresponding features were separated by Metric Learning such that

$$\left\|\hat{\phi}_i - \hat{\phi}_j\right\|_2 \gg \sigma,\ \forall a\left(\hat{\phi}_i\right) \ne a\left(\hat{\phi}_j\right) \tag{8}$$

where $a\left(\hat{\phi}_i\right)$ is the age of the face encoded by $\hat{\phi}_i$, we have that

$$K\left(\hat{\phi}_i, \hat{\phi}_j\right) \approx 0,\ \forall a\left(\hat{\phi}_i\right) \ne a\left(\hat{\phi}_j\right). \tag{9}$$

Thus, following Eqs. (8) and (9), for a test sample $\hat{\phi}$ related to an age $a\left(\hat{\phi}\right)$

$$\begin{aligned} f\left(\hat{\phi}\right) &= \sum_i \alpha_i K\left(\hat{\phi}, \hat{\phi}_i\right) + b \\ &= \sum_{a\left(\hat{\phi}_i\right) = a\left(\hat{\phi}\right)} \alpha_i K\left(\hat{\phi}, \hat{\phi}_i\right) + b \end{aligned} \tag{10}$$

implying that the test sample $\hat{\phi}$ only interacts (i.e. $K\left(\hat{\phi},\hat{\phi}_i\right)\neq 0$) with support vectors $\hat{\phi}_i$ such that $a\left(\hat{\phi}_i\right)=a\left(\hat{\phi}\right)$. Thus, the regression error would be similar to the training error, that is the optimal error one can strive for.

4. Experimental results

The proposed age regression scheme was experimentally evaluated by applying it to contemporary state-of-the-art datasets. For that we studied datasets of face images with corresponding ages, such as the MORPH-II [22] and FG-Net [20], rather than datasets consisting of age classes [6,38,39]. We were unable to utilize the Chalearn dataset [36], as it is based on manual age annotation, without an accuracy estimate. Moreover, the accuracy of the proposed scheme, when applied to the MORPH-II dataset, is similar to the accuracy of human-based age estimation, and cannot be used as a baseline for accurate age estimation schemes such as ours.

The MORPH-II Database [22] consists of 55,134 images of 13,000 subjects and their ages. All images were taken in controlled settings with similar illumination conditions and subject pose. The images were preprocessed and augmented by first computing and subtracting an average image, and then horizontally flipping the images and applying horizontal and vertical translations of ±1 pixels. This augmentation yields close to one million images, making it suitable for full CNN training.

Due to the small number of images in the FG-Net dataset, making it unsuitable for training CNNs from scratch, we followed the approach of Demontis et al. [28] and Ni et al. [30], that trained their age estimation schemes on the FRGC [29] and MORPH-I datasets, respectively, and evaluated their accuracy using the FG-Net dataset.

We first trained a CNN using a Softmax loss following the approach detailed in Section 3.2, using the MORPH-II dataset. The CNN was refined by minimizing the L_2 loss of the estimated age. Its FC6 layer was used as a face descriptor ϕ_i, to which we applied Metric Learning and SVR as detailed in Section 3, where $\hat{d}=4$ was the dimension of the projected vector $\hat{\phi}$. We used 60% of the data for training, 20% for validation and 20% for testing. The same training data partitions were used for identity and age classification.

The proposed CNN was trained using SGD with a batch size of $n=200$ images, weight decay of 0.0002 and momentum of 0.9. The weights in each layer were initiated by a zero-mean Gaussian distribution with standard deviation 0.01 and the biases set to zero. An equal learning rate was used in all layers, which was initialized to 0.01 and was automatically decreased along epochs following the decrease rate of the training error, down to 0.0001. In the spatial normalization layer, the hyper parameters were set to $\kappa=2, n=5, \alpha=10^{-4}$ and $\beta=0.75$, while the dropout probability was set to 0.5.

We also used the FC6 layer of the VGG-Face CNN [19], that is publicly available,[1] as a transfer learning-based representation. The VGG-Face was trained using a dataset of 2.6M face images (before augmentation), and achieves state-of-the-art accuracy on the LFW dataset [19]. The gist of this approach, is to utilize transfer learning, as in some applications such as age estimation, it is practically impossible to train an ultra-deep net such as VGG-Face from scratch, due to the lack of training data and the required computational resources. In contrast, the refinement step can be trained using significantly smaller training sets. For that we implemented an L_2 regression loss layer.

The Metric Learning metric $\mathbf{W}$ was trained using SGD with a learning rate of $\gamma=0.1$ and a bias $\gamma_b=12$, where $\mathbf{W}$ was initialized by the $\hat{d}$ leading PCA vectors of the learning set of the image descriptors ϕ_i, whitened by the inverse of the corresponding eigenvalues. The Kernel-SVR was applied using LibSVM [40], grid search and cross-validation to optimize the parameters of the RBF kernel and soft margins.

[1] http://www.robots.ox.ac.uk/vgg/software/vgg_face/.

Table 1
Age estimation results for the MORPH-II datasets for contemporary state-of-the-art schemes. We report the MAE and CS accuracy metrics.

Method	MAE	CS(5) %
AGES [41]	8.07	46%
RED-SVM [42]	6.49	49.5%
OHRank [8]	6.07	56.3%
Han et al. [43]	4.2	72.4%
Wang [11]	3.81 (Caucasian only)	–
Wang [33]	4.77	
Yi [14]	3.66	–
DAR	3.16	80.35%
DAC [13]	3.14	80.67%
VGG-Face [19] + L_2 loss	3.44	76.31%
DeepAge	**2.87**	**84.18%**

As figures of merits for evaluating the accuracy of the proposed scheme, we report the Mean Absolute Errors (MAE)

$$MAE=\frac{1}{N}\sum_i|\hat{a}_i-a_i|, \tag{11}$$

where a_i is the ground truth age, $\hat{a}_i$ is the estimated age, and N is the number of test images, and the Cumulative Score (CS)

$$CS(j)=\frac{N\,(e\leq j)}{N}, \tag{12}$$

where $N\,(e\leq\varepsilon)$ is the number of test images whose absolute estimation error is less than ε. We report $CS(5)$ following contemporary works.

4.1. The MORPH-II dataset

We compared the proposed scheme to prior state-of-the-art results reported for the MORPH-II dataset, and also implemented the CNN-based approach of Hassner and Levi [13], by training a classification network over the set of 55 age classes corresponding to one of the ages in the MORPH-II dataset. We denote this approach Deep Age Classification (DAC), and also implemented a deep regression network using the L_2 loss of the estimated age, that is denoted as Deep Age Regression (DAR). We also used the VGG-Face CNN [19] as a Face Classification Network (FCN) used to train an Age Regression Network (ARN), by replacing its Softmax loss of the FCN with a L_2 regression loss and retraining the FC layers. Trying to apply Metric Learning and SVR to the FC layer $\phi_i^{VGG}\in\mathbb{R}^{4096}$, used as a descriptor, resulted in inferior accuracy.

The age estimation accuracy results are reported in Table 1, where it follows that the proposed scheme outperforms the previous results, as well as the DAR, DAC and VGG-Face-based Deep Learning approaches. We note that all Deep Learning-based schemes outperform previous (shallow) state-of-the-art approaches.

We study the sensitivity of the MAE and CS metrics to the choice of the FC layer used for the representation ϕ_i, and the corresponding dimensionality reduction $\hat{d}$ in Table 2, by comparing the results of applying the proposed Metric Learning (ML) scheme (Section 3.3) to PCA. The projection $\mathbf{W}$ and SVR were trained separately for each of the different layers, and ML dimensions $\hat{d}$. It follows that using the FC6 layer outperforms the FC7 layer, while both significantly outperform the PCA a dimensionality reduction scheme. Setting $\hat{d}=4$ results in the highest accuracy, although other choices of $\hat{d}$ yield similar accuracies that significantly outperforms the ones achieved by applying PCA. We attribute the improved accuracy to the supervised training of the ML-based dimensionality reduction, compared to the unsupervised PCA.

This is further exemplified in Fig. 5 where we depict the Cumulative Score (CS) for using FC6 and FC7 for different PCA and ML dimensionalities. It follows that as in Table 2 (for CS(5)) the proposed ML approach improves on the commonly used PCA.

Table 2
Age estimation accuracy of the DeepAge scheme when applied to the MORPH-II dataset. We report the results using FC6 and FC7 for representation learning. Dimensionality reduction is applied via PCA and Metric Learning for varying dimensionalities.

FC 6				
$\hat{d}$	DL		PCA	
	MAE	CS(5)%	MAE	CS(5)%
2	2.89	83.98%	5.67	42.46%
3	2.88	84.06%	5.68	42.37%
4	**2.87**	84.18%	6.09	38.28%
8	2.89	83.93%	5.74	43.15%
16	2.91	**84.26%**	5.92	41.12%
32	2.93	83.76%	5.78	43.18%
128	3.12	81.24%	5.81	43.86%
FC 7				
$\hat{d}$	DL		PCA	
	MAE	CS(5)%	MAE	CS(5)%
2	3.03	81.36%	4.11	64.67%
3	2.99	82.02%	4.20	62.99%
4	2.98	82.33%	4.20	63.24%
8	2.93	83.06%	4.19	63.23%
16	3.32	77.66%	3.79	69.59%
32	3.23	79.29%	3.83	69.44%
128	3.05	81.72%	4.70	56.08%

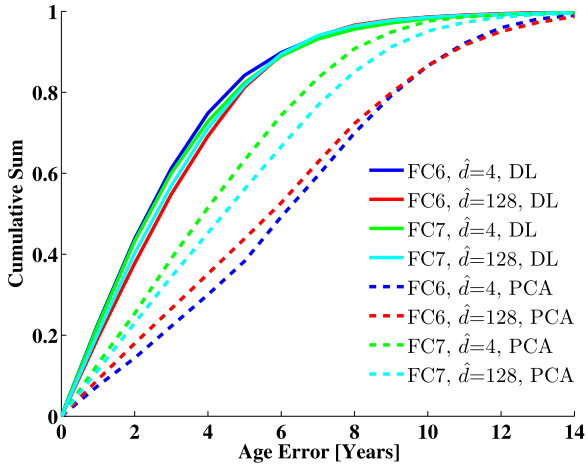

Fig. 5. MORPH-II age estimation accuracy in terms of the Cumulative Score (CS). We present the results of using FC6 and FC7 for varying dimensionalities $\hat{d}$ for PCA and Metric Learning.

The effectivity of the Metric Learning scheme is also shown in Fig. 6 where we represent each face using its two leading embedding coordinates $\{\phi_1, \phi_2\}$, and the color encodes the age. It follows that the proposed ML scheme forms a one-dimensional banana-shaped manifold in the two-dimensional $\{\phi_1, \phi_2\}$ domain, where the age parametrization is the principal non-linear axis of the data.

Fig. 7 depicts age estimation results, where we show the lowest and highest age estimation errors, in the upper and lower image rows, respectively. In particular, some of the erroneous estimates, in Fig. 7e–g seem visually justified, and might relate to errors in the MORPH-II dataset.

4.2. FG-Net dataset

The proposed scheme was also evaluated using the FG-Net Aging Database [20] that consists of only 1002 images of 82 different subjects. A comprehensive review of the age estimation schemes which were evaluated using this dataset, and their corresponding results can be found in [44]. Most FG-Net results were computed using the Leave One Person Out (LOPO) approach, where for each of the 82 subjects in the database, an age estimator is trained using images of the remaining 81 subjects and the results are averaged over all 82 subjects. Using such a small training set to train the proposed scheme from scratch, would lead to overfitting. Hence, we first compare our results to works, where the age estimator was trained using a *different*, larger face dataset and evaluated using the FG-NET dataset [28,30]. For that we applied the CNN used in Section 4.1 that was trained using the MORPH-II dataset, and the results are reported in Table 3, where we compare to Demontis et al. [28] and Ni et al. [30].

As neither the FCN and ARN networks could be trained form scratch using the FG-NET dataset due to its small size, only the Kernel-SVR phase was retrained using subsets of the FG-NET images. First, we trained the Kernel-SVR using 200 randomly drawn FG-NET images, and

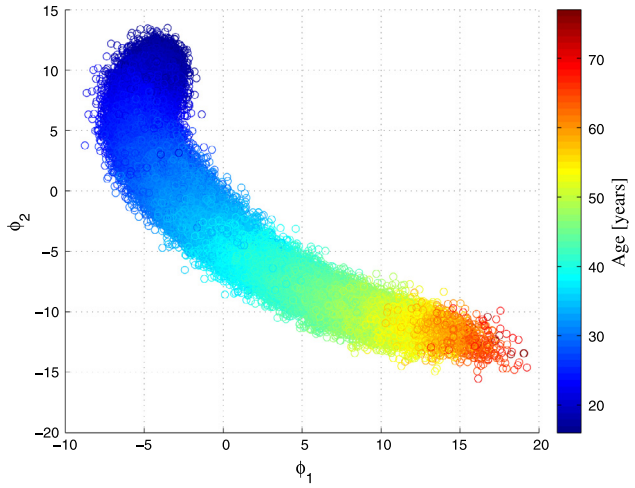

Fig. 6. Embedding of the MORPH-II dataset. We plot the coordinates of the two leading embedding vectors. The color corresponds to the subjects age. Please review this plot in color.

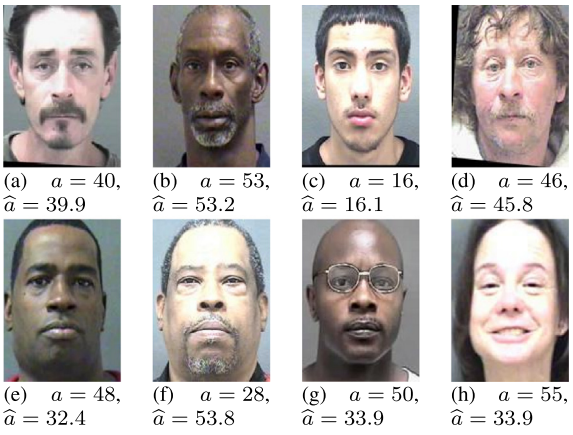

Fig. 7. Age estimation results, where a and $\hat{a}$ are the groundtruth and estimated ages, respectively. The upper row depicts successful estimates, while the lower one relates to significant age estimation errors.

Table 3
Age estimation results of the FG-Net datasets of schemes that were trained using other datasets.

Method	MAE
Raw pixels [28]	12.58
PCA [28]	12.48
LDA [28]	13.28
Web dataset [30]	9.49
MORPH-I [30]	10.37
DeepAge	**9.10**

Table 4
Age estimation results for the FG-Net using a training set of 200 FG-Net images. Rows #1−#2 report the DeepAge results trained using MORPH-II. Rows #3−#4 report the results of retraining the SVR using the FG-Net, while Rows #5-#6 show the results of applying DeepAge using a linear regression trained using the FG-Net. Row #7 shows the results of refining the VGG-FACE ARN using the MORPH-II dataset, while row #8 depicts the results of refining the FC layer of the VGG-FACE ARN.

Parameters	MAE	CS(5) [%]
DeepAge FC6	9.10	20.00
DeepAge FC7	10.52	10.00
DeepAge FC6 refinement	**7.08**	43.33
DeepAge FC7 refinement	7.88	43.33
DeepAge FC6 + L_2 loss refinement	8.57	36.82
DeepAge FC7 + L_2 loss refinement	9.41	32.73
VGG-Face [19] + L_2 loss	9.68	36.67
VGG-Face [19] + L_2 loss refinement	7.43	**44.87**

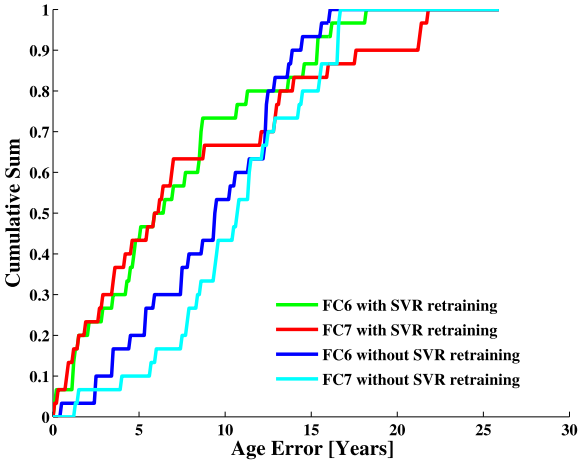

Fig. 8. Age estimation accuracy of the proposed DeepAge scheme for the FG-NET dataset. We show the Cumulative Score (CS) while using FC6 and FC7 for representation learning, with and without SVR refinement using a subset of the FG-Net dataset.

second, we applied the LOPO evaluation and retrained the Kernel-SVR 82 times, using $\backsim$ 1 K training images.

We compared the results of training using 200 random images against multiple schemes. First, we applied the VGG-Face-based regression CNN used in Section 4.1, and refined using the MORPH-II. Second, we refined the FC layer of the VGG-Face CNN using the same 200 FG-NET images used to refine the proposed scheme. Last, in order to assess the upside of using SVR in contrast to linear L_2 regression, we reapplied the DeepAge scheme such that the SVR was replaced by linear regression, applied to the output of the metric learning phase.

The results are reported in Table 4 where it follows that the retraining improved the age estimation accuracy significantly for both ARN and VGG-Face CNNs, and using the SVR improved the accuracy, compared to using linear regression. The VGG-Face-based regression CNN provides a more robust, but less accurate solution, as it improves the $CS(5)$ score, but not the MAE. This implies that it is less prone to outliers, and provides similar estimation accuracy. We attribute that to the significantly larger training set of the VGG-Face (2.6M images) compared to the ARN trained using only ∼36K images.

Similar results are presented in Fig. 8 that depicts the Cumulative Score error measure for varying refinement strategies, and it follows that the SVR retraining using a small subset of the FG-Net dataset improves the accuracy. Thus, it seems that domain adaptation can be achieved in different phases of the scheme, and in particular, the Kernel-SVR layer might be considered a good choice, as it requires a smaller training set than the Deep Learning layers.

The results of the LOPO evaluation are reported in Table 5, where we compare against the schemes in the survey reported by Panis and Lanitis [44], showing the highest accuracy. The proposed DeepAge is shown to outperform previous results by achieving an MAE of 3.01. This accuracy significantly outperforms the accuracy of the DeepAge scheme in Table 4, that was trained using 200 images, in contrast to the $\backsim$ 1 K training images used in the LOPO evaluation. We note that the previous FG-NET schemes were trained from scratch using only the FG-NET images, while the DeepAge was initially trained using $\backsim$36 K MORPH-II images.

Table 5
Age estimation results for the FG-Net dataset using the Leave One Person Out (LOPO) evaluation. The results of the previous schemes were reported by Panis and Lanitis [44].

Parameters	MAE
Kilinc2013 [45]	5.05
Chang2011 [8]	4.48
Chao2013 [46]	4.38
Hong2013 [47]	4.18
El Dib2010 [48]	3.17
DeepAge	**3.01**

5. Conclusions and future work

In this work we presented a computational approach for face-based age estimation. We propose to utilize CNNs trained using large scale identification tasks, and refined using smaller age regression training sets for Representation Learning. In that we aim to overcome the need for large training sets needed for training high performing ultra-deep CNNs. Age supervised Metric Learning is applied for dimensionality reduction and Kernel-SVR was used for regression. We also show how to utilize a pre-trained ultra-deep CNN, such as VGG-Face, for Representation Learning, avoiding the need to train task specific CNNs. The proposed scheme was applied to the MORPH-II and FG-Net datasets and was shown to compare favorably with contemporary state-of-the-art approaches. In particular, we show that domain adaptation that is essential for analyzing small-scale datasets such as the FG-Net, and can be achieved by retraining the SVR layer, rather than the CNN.

References

[1] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.
[2] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014, pp. 1701–1708.
[3] E. Ramón-Balmaseda, J. Lorenzo-Navarro, M. Castrillón-Santana, Gender classification in large databases, in: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Springer, 2012, pp. 74–81.
[4] S. Mahpod, Y. Keller, Kinship verification using multiview hybrid distance learning, Comput. Vis. Image Underst. 167 (2018) 28–36.
[5] E. Dahan, Y. Keller, S. Mahpod, Kin-verification model on fiw dataset using multi-set learning and local features, in: Proceedings of the 2017 Workshop on Recognizing Families in the Wild, in: RFIW '17, ACM, New York, NY, USA, 2017, pp. 31–35.
[6] E. Eidinger, R. Enbar, T. Hassner, Age and gender estimation of unfiltered faces, Inf. Forensics Secur., IEEE Trans. 9 (12) (2014) 2170–2179.
[7] G. Guo, G. Mu, Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 657–664.
[8] K.-Y. Chang, C.-S. Chen, Y.-P. Hung, Ordinal hyperplanes ranker with cost sensitivities for age estimation, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 585–592.
[9] D. Cao, Z. Lei, Z. Zhang, J. Feng, S. Li, Human age estimation using ranking svm, in: W.-S. Zheng, Z. Sun, Y. Wang, X. Chen, P. Yuen, J. Lai (Eds.), Biometric Recognition, in: Lecture Notes in Computer Science, vol. 7701, Springer Berlin Heidelberg, 2012, pp. 324–331.
[10] W.-L. Chao, J.-Z. Liu, J.-J. Ding, Facial age estimation based on label-sensitive learning and age-oriented regression, Pattern Recognit. 46 (3) (2013) 628–641.
[11] X. Wang, C. Kambhamettu, Age estimation via unsupervised neural networks, in: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, Vol. 1, 2015, pp. 1–6.

[12] K. Chen, S. Gong, T. Xiang, C. Loy, Cumulative attribute space for age and crowd density estimation, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2467–2474.
[13] G. Levi, T. Hassncer, Age and gender classification using convolutional neural networks, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on, 2015, pp. 34–42.
[14] D. Yi, Z. Lei, S. Z.L.i, Age estimation by multi-scale convolutional network, in: Computer Vision–ACCV 2014, Springer, 2015, pp. 144–158.
[15] X. Xiong, F. De la Torre, Supervised descent method and its applications to face alignment, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 532–539.
[16] S. Yan, M. Liu, T. Huang, Extracting age information from local spatially flexible patches, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, pp. 737–740.
[17] S.E. Choi, Y.J. Lee, S.J. Lee, K.R. Park, J. Kim, Age estimation using a hierarchical classifier based on global and local facial features, Pattern Recognit. 44 (6) (2011) 1262–1281.
[18] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.
[19] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, 2015.
[20] T. Cootes, A. Lanitis, The fg-net aging database, 2002. Available online at http://www-prima.inrialpes.fr/FGnet/.
[21] G. Guo, G. Mu, Human age estimation: What is the influence across race and gender?, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, pp. 71–78.
[22] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006, pp. 341–345.
[23] H. Han, C. Otto, A. Jain, Age estimation from face images: Human vs. machine performance, in: International Conference on Biometrics (ICB), 2013, pp. 1–8.
[24] A. Lanitis, C. Draganova, C. Christodoulou, Comparing different classifiers for automatic age estimation, IEEE Trans. Syst. Man Cybern. B 34 (1) (2004) 621–628.
[25] Y. Fu, T. Huang, Human age estimation with regression on discriminative aging manifold, Multimedia, IEEE Trans. 10 (4) (2008) 578–584.
[26] Y. Fu, Y. Xu, T. Huang, Estimating human age by manifold analysis of face pictures and regression on aging features, in: IEEE International Conference on Multimedia and Expo, 2007, pp. 1383–1386.
[27] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, T. Huang, Regression from patch-kernel, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.
[28] A. Demontis, B. Biggio, G. Fumera, F. Roli, Super-sparse regression for fast age estimation from faces at test time, in: Image Analysis and Processing—ICIAP 2015, Springer, 2015, pp. 551–562.
[29] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, 2005, pp. 947–954.
[30] B. Ni, Z. Song, S. Yan, Web image and video mining towards universal and robust age estimator, Multimedia, IEEE Trans. 13 (6) (2011) 1217–1229.
[31] G. Guo, Y. Fu, C. Dyer, T. Huang, Image-based human age estimation by manifold learning and locally adjusted robust regression, IEEE Trans. Image Process. 17 (7) (2008) 1178–1188.
[32] G. Guo, G. Mu, Y. Fu, C. Dyer, T. Huang, A study on automatic age estimation using a large database, in: IEEE 12th International Conference on Computer Vision, 2009, pp. 1986–1991.
[33] X. Wang, R. Guo, C. Kambhamettu, Deeply-learned feature for age estimation, in: 2015 IEEE Winter Conference on Applications of Computer Vision, 2015, pp. 534–541.
[34] C.B. Choy, J. Gwak, S. Savarese, M. Chandraker, Universal correspondence network, in: Advances in Neural Information Processing Systems, Vol. 29, 2016.
[35] K.M. Yi, E. Trulls, V. Lepetit, P. Fua, Lift: Learned invariant feature transform, in: Proceedings of the European Conference on Computer Vision, 2016.
[36] H.J. Escalante, V. Ponce-Lpez, J. Wan, M.A. Riegler, B. Chen, A. Claps, S. Escalera, I. Guyon, X. Bar, P. Halvorsen, H. Mller, M. Larson, Chalearn joint contest on multimedia challenges beyond visual analysis: An overview, in: 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 67–73.
[37] A. Krizhevsky, I. Sutskever, G. E.H.inton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[38] K. Ueki, T. Hayashida, T. Kobayashi, Subspace-based age-group classification using facial images under various lighting conditions, in: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006, pp. 6 pp.–48.
[39] A. Gallagher, T. Chen, Understanding images of groups of people, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 256–263.
[40] C.-C. Chang, C.-J. Lin, Libsvm: A library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (2011) 27:1–27:27.
[41] 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.
[42] K.-Y. Chang, C.-S. Chen, Y.-P. Hung, A ranking approach for human ages estimation based on face images, in: Pattern Recognition (ICPR), 2010 20th International Conference on, 2010, pp. 3396–3399.
[43] H. Han, C. Otto, A. Jain, Age estimation from face images: Human vs. machine performance, in: Biometrics (ICB), 2013 International Conference on, 2013, pp. 1–8.
[44] G. Panis, A. Lanitis, An Overview of Research Activities in Facial Age Estimation Using the FG-NET Aging Database, Springer International Publishing, Cham, 2015, pp. 737–750.
[45] M. Kilinc, Y.S. Akgul, Automatic human age estimation using overlapped age groups, in: G. Csurka, M. Kraus, R.S. Laramee, P. Richard, J. Braz (Eds.), Computer Vision, Imaging and Computer Graphics. Theory and Application, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 313–325.
[46] W.-L. Chao, J.-Z. Liu, J.-J. Ding, Facial age estimation based on label-sensitive learning and age-oriented regression, Pattern Recognit. 46 (3) (2013) 628–641.
[47] L. Hong, D. Wen, C. Fang, X. Ding, A new biologically inspired active appearance model for face age estimation by using local ordinal ranking, in: Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, in: ICIMCS '13, ACM, New York, NY, USA, 2013, pp. 327–330.
[48] M.Y.E. Dib, M. El-Saban, Human age estimation using enhanced bio-inspired features (ebif), in: 2010 IEEE International Conference on Image Processing, 2010, pp. 1589–1592.