Accepted in Pattern Recognition Letters

Pattern Recognition Letters
journal homepage: www.elsevier.com

Are you eligible? Predicting adulthood from face images via class specific mean autoencoder

Maneet Singh, Shruti Nagpal, Mayank Vatsa**, Richa Singh

IIIT-Delhi, New Delhi, 110020, India

Article history:
Received 15 March 2017

Keywords: Adulthood Prediction, Supervised Autoencoder, Face Analysis

ABSTRACT

Predicting if a person is an *adult* or a *minor* has several applications such as inspecting underage driving, preventing purchase of alcohol and tobacco by minors, and granting restricted access. The challenging nature of this problem arises due to the complex and unique physiological changes that are observed with age progression. This paper presents a novel deep learning based formulation, termed as Class Specific Mean Autoencoder, to learn the intra-class similarity and extract *class-specific* features. We propose that the feature of a particular class if brought similar/closer to the *mean* feature of that class can help in learning class-specific representations. The proposed formulation is applied for the task of *adulthood* classification which predicts whether the given face image is of an adult or not. Experiments are performed on two large databases and the results show that the proposed algorithm yields higher classification accuracy compared to existing algorithms and a Commercial-Off-The-Shelf system.

1. Introduction

Human aging is a complex process and brings with it behavioral and physiological changes. One associates maturity and mental growth with the behavioral changes that occur with time. Age is also often used as a means of access control, physically as well as virtually, to keep younger minds away from activities and content they are not deemed ready for. A threshold age, known as the *age of majority*, is defined by most states to universalize the concept of an individual being physically and mentally ready to assume control for their actions and decisions. However, there are different age limits prescribed by individual state and federal governments for different activities. For instance, in the USA, the legal age for smoking is 18 years while age limit for voting and drinking is 21 years.

The physiological and behavioral effects of aging vary for every individual and are a function of several parameters such as health, living style, environmental conditions, and gender. Therefore, it is challenging to accurately estimate the age of a person. As can be seen from Fig. 1, it becomes difficult to predict the age of individuals just immediately below, or above the age of majority (e.g. 18 years). Among the currently avail-

(a) Below the age of majority

(b) Above the age of majority

Fig. 1. Sample images from the FG-NET Aging dataset (Panis et al., 2016). (a) shows images of individuals below the age of majority, and (b) shows sample individuals at the age of majority. These examples illustrate the challenging nature of *adulthood classification*.

**Corresponding author: Tel.: +91-9654653404; fax: +91-11-26907410;
e-mail: mayank@iiitd.ac.in (Mayank Vatsa)

able non-intrusive biometrics, the face *changes* significantly with age and is therefore a preferred modality for estimating the age of a person. As of now, at different checkpoints, an officer or a designated person in-charge estimates the age of a person by visually observing an individual. In cases where visual inspection becomes difficult, she/he asks for an identification (ID) card as the proof of age. However, research has shown that both these measures used for age estimation and verification are prone to errors (Martinez et al., 2007; Ferguson and Wilkinson, 2017). With easy availability of tampered or fake ID cards, several youngsters use these cards to mis-represent their identity. Recently, in a survey at Harvard University (Wechsler et al., 2002), it was found that 18% underage students obtained alcohol using fake ID cards. Inspired by these observations and motivated by the increasing use of technology and automation in our day-to-day life, this research aims to automate the process of classifying an individual as an adult or not. For a given face image, the proposed algorithm aims to predict if the person has attained the age of majority or not. Such a system could be deployed at multiple places having age based restricted access; for instance, voting centers, driving license centers, and traffic check posts to scan for minors, restaurants and bars to prevent under-age alcohol consumption, cinema halls to enable restricted access, or around tobacco selling vending machines. Apart from the above mentioned applications, such systems can also be deployed virtually, where access is granted based on the age; for example, in online poker rooms.

1.1. Literature Review: Age Estimation

In literature, researchers have focused on estimating the age of an individual either by classification or by regression, and via hybrid methods which are a combination of both (Fu et al., 2010). On considering it as a classification task, the problem is formulated as a n-class problem, where each age label is an independent class, or the age range is divided into groups, such as child, young adult, adult, and old. On the other hand, in case of regression, the age value to be estimated is considered as a series of sequential numbers. Guo et al. (2008) have proposed a robust method to extract facial features using manifold learning and a novel method, Locally Adjusted Robust Regressor (LARR) is presented to predict the age, where support vector regressors are explored on the learned manifold. Luu et al. (2009) have proposed an age estimation hybrid technique based on Active Appearance Models (AAMs) and Support Vector Regression (SVR), wherein age-group specific models are utilized to perform more accurate prediction. AAMs are used to extract discriminative features from the face images based on shape as well as textural changes, while SVR is used to perform age estimation on the extracted features. Pontes et al. (2016) utilized several hand-crafted features such as AAM, Local Binary Patterns, Gabor Wavelets, and Local Phase Quantization, along with Support Vector Machines (SVMs) and SVRs for performing age estimation. In order to avoid overfitting while performing the task of age classification, Eidinger et al. (2014) have proposed a novel approach for training SVMs using dropout. The authors show that the modified model improves the classification accuracy for the defined problem (8 age groups) as compared to the existing benchmark results. Recently, Geng et al. (2013) have proposed to formulate the problem of age estimation by using a label distribution for a given sample rather than a single label. The label distribution represents the proportion of each age label, which is motivated by the fact that neighboring ages are highly correlated in nature, and hence cannot be described by a single age value. Han et al. (2013) studied the performance of humans versus machines for age estimation and established machines' performance to supersede that of humans on the FG-Net Aging database (Panis et al., 2016). Inspired by the promising performance of deep learning architectures, Levi and Hassner (2015) propose a lean shallow network using Convolutional Neural Networks (CNNs) for attribute prediction of unconstrained images. Recently, Li et al. (2017) and Xing et al. (2017) have also demonstrated superior performance of deep learning models for performing age estimation on face images.

Fig. 2. Demonstrating the intra-class variations in the two classes: first row represents images belonging to *minor* class and second row corresponds to *adult* class.

In 2015, ChaLearn Looking At People (LAP) also organized a challenge as part of ICCV'15 (Escalera et al., 2015) to perform *apparent* age estimation. Apparent age is defined as the age perceived by human beings based on the visual appearance of an individual. The dataset was created using a Facebook application to vote for the perceived age from a given image. Rothe et al. (2015) obtained the best performance in the given challenge, wherein the authors presented a CNN based model. The VGG-16 models (Simonyan and Zisserman, 2014) pre-trained on Image-Net dataset (Russakovsky et al., 2014) are fine-tuned using the proposed IMDB-Wiki datasets for the task of apparent age estimation, and a single neuron gives the output, a whole number from [0, 100]. However, it is important to note that the competition is aimed at predicting the apparent or perceived age of face images, which is less relevant for monitoring cases of restricted access, as compared to the real age.

1.2. Research Contributions

In this research, we propose a deep learning based novel representation learning algorithm to determine whether the given input face image is above the age of majority or not, i.e., the image corresponds to an adult or a minor. A supervised deep learning algorithm is presented which reduces the intra-class

variations at the time of feature learning. Thus, the three-fold contributions of this research are:

- propose the formulation of a deep learning architecture termed as the *Class Specific Mean Autoencoder*, which uses the class information of a given sample at the time of training to learn the intra-class similarity and extract similar features for samples belonging to the same class,
- present the Multi-Resolution Face Dataset (MRFD) which contains images pertaining to 317 subjects (each having at least 12 images), out of which 307 subjects are below the threshold age of 18 years. MRFD is created due to the lack of an existing database containing images of subjects below the age of 18 years and will be made publicly available to the research community,
- demonstrate results on the Multi-Resolution Face Dataset as part of a large combined face database of more than 13,000 images from multiple ethnicities. Results are also demonstrated on MORPH Album II database (Ricanek Jr. and Tesafaye, 2006) containing more than 55,000 face images.

The remainder of this paper is organized as follows: Section 2 provides the detailed description of the proposed algorithm. Section 3 presents the datasets used, along with the experimental protocols. The results and observations are presented in Section 4, followed by the conclusions of this work.

2. Proposed Algorithm

Deep learning architectures have been used in literature to address a large variety of tasks (Lecun et al., 2015). Specifically, recent models such as the FaceNet (Schroff et al., 2015), VGG-Face (Parkhi et al., 2015), and DeepFace (Taigman et al., 2014) have shown high performance for the task of face recognition. Models have been developed to perform automated face detection and alignment as well (Li et al., 2015; Farfade et al., 2015; Chen et al., 2014). In this work, the task of adulthood classification is addressed using a deep learning framework. As shown in Fig. 2, the intra-class variability in both classes, minor and adult, is high. Analyzing the mean image of individual classes (as shown in Fig. 3) shows that the mean images of the two classes are significantly different. Based on this observation, it is our hypothesis that projecting the image features closer to the class mean can assist in learning class specific discriminative features. Therefore, in this work, we propose Class Specific Mean Autoencoder, which learns features such that the representations of the samples corresponding to a class are similar to the mean representation of the same class. Before elaborating upon the proposed model, the following subsection presents some preliminaries.

2.1. Preliminaries: Supervised Autoencoders

Several researchers have proposed modifications to the traditional autoencoder architecture. Table 1 provides a summary

Fig. 3. Mean face images obtained from the images corresponding to the two age groups. The left image corresponds to the mean image of individuals below the age of majority, while the right image corresponds to the mean image of individuals above the age of majority.

of these architectures. Most of these are unsupervised in nature, however, researchers have proposed supervised architectures that leverage the availability of labeled data as well. In this section, we briefly present the original formulation of autoencoder followed by discussing the existing supervised architectures.

For a given input x, the loss function of a single layer traditional autoencoder (Hinton and Salakhutdinov, 2006) is given as follows:

$$\underset{\mathbf{W}_e,\, \mathbf{W}_d}{\arg\min} \|x - \mathbf{W}_d \phi(\mathbf{W}_e x)\|_2^2 \tag{1}$$

where, $\mathbf{W}_e$ and $\mathbf{W}_d$ are the respective encoding and decoding weights of the autoencoder, and ϕ corresponds to an activation function, generally incorporated for introducing nonlinearity in the model. Common examples of activation functions are *sigmoid* and *tanh*. An autoencoder learns features ($f_x = \phi(\mathbf{W}_e x)$) of the given input x, such that the error between the original sample and it's reconstruction ($\mathbf{W}_d f_x$) is minimized. For a k layered autoencoder, having encoding weights as $\mathbf{W}_e^1, \mathbf{W}_e^2, ..., \mathbf{W}_e^k$, and decoding weights as $\mathbf{W}_d^1, \mathbf{W}_d^2, ..., \mathbf{W}_d^k$, the loss function of Eq. 1 is modified as follows:

$$\underset{\mathbf{w}_e^1,...,\mathbf{w}_e^k,\mathbf{w}_d^1,...,\mathbf{w}_d^k}{\arg\min} \|x - b \circ a\,(x)\|_2^2 \tag{2}$$

where, $a(x) = \phi(\mathbf{W}_e^k(\phi(\mathbf{W}_e^{k-1} \dots (\phi(\mathbf{W}_e^1 x)))))$ refers to the encoding function, and $b\,(x) = \mathbf{W}_d^1(\mathbf{W}_d^2 \dots (\mathbf{W}_d^k x))$ corresponds to the decoding function. The first and the last layers correspond to the input and output layers respectively, while the remaining layers are often termed as the hidden layers.

In literature, researchers have incorporated class information in the traditional formulation of an autoencoder in order to facilitate supervision. Gao et al. (2015) modify the denoising autoencoder (Vincent et al., 2010) to learn supervised image representations in order to optimize the identification performance. At the time of training, for a given subject, the probe image is the input to the autoencoder (analogous to the noisy input), and the gallery image of the subject (analogous to the clean image) is the target image used for computing the reconstruction error, as in the case of a denoising autoencoder. A similarity preservation term is added to the loss function such that the samples belonging to the same class have a similar representation. Given probe and gallery images of class i, each probe image is represented using x_{ni} and its corresponding gallery images are represented using x_i. The loss function for the supervised autoen-

Table 1. Brief literature review of autoencoder based formulations.

Authors	Approach	Supervised
Vincent et al. (2010)	Stacked Denoising Autoencoder (SDAE): Noise is added to the input data such that the learned representation is robust.	No
Ng (2011)	Incorporated ℓ_1 norm in the loss function of the autoencoder to introduce sparsity in the learned features.	No
Rifai et al. (2011b)	Contractive Autoencoder (CAE): Input space is localised by adding a penalty term which is the Jacobian of the input with respect to the hidden layer.	No
Rifai et al. (2011a)	Higher order Contractive autoencoder: CAE + Hessian of the output wrt the input.	No
Zheng et al. (2014)	Contrastive autoencoder: A term to reduce the intra-class variations between the learned representation of samples belonging to the same class is added at the final layer.	Yes
Wang et al. (2014)	Generalised Autoencoder: SDAE is modified such that the representation incorporates the structure of the dataspace as well.	No
Zhang et al. (2015)	Stacked Multichannel Autoencoder: The gap between synthetic data and real data is reduced by learning a mapping between the two.	No
Gao et al. (2015)	Inspired from SDAE, an identification specific model is proposed, where the probe image is treated as the noisy input while the gallery images are treated as the clean input.	Yes
Zhuang et al. (2015)	A two layer model is proposed wherein, a representation is learned in the first layer, and the class label is encoded in the second layer.	Yes
Majumdar et al. (2017)	A joint sparsity (using $\ell_{2,1}$) promoting supervision penalty term is added to the loss function of SDAE.	Yes
Meng et al. (2017)	A relational term, which aims to model the relationship between the input data is added to the loss function.	No
Proposed (2018)	Class Specific Mean Autoencoder: utilizes class mean to learn discriminative features.	Yes

coder is as follows:

$$\begin{aligned}\underset{\substack{\mathbf{w}_e^1,\ldots,\mathbf{w}_e^k,\\ \mathbf{w}_d^1,\ldots,\mathbf{w}_d^k}}{\arg\min}\ \frac{1}{N}\sum_i \Big(\|x_i - b\circ a(x_{ni})\|_2^2 + \lambda\,\|a(x_i)-a(x_{ni})\|_2^2\Big)\\ +\alpha\Big(KL(\rho_x\|\rho_o) + KL(\rho_{x_n}\|\rho_o)\Big)\\ where, \rho_x = \frac{1}{N}\sum_i \frac{1}{2}\Big(a(x_i)+1\Big)\quad and\\ \rho_{x_{ni}} = \frac{1}{N}\sum_i \frac{1}{2}\Big(a(x_{ni})+1\Big)\end{aligned} \tag{3}$$

here, the first term corresponds to the reconstruction error, second is the similarity preservation term, and the remaining two terms correspond to the Kullback Leibler (KL) divergence (Kullback and Leibler, 1951) to introduce sparsity in the hidden layers.

Contrastive Autoencoder (CsAE) proposed by Zheng et al. (2014), is another variant of supervised autoencoder which uses the class label information during training. The loss function of the model is the difference between the output of two sub-autoencoders trained simultaneously on samples belonging to the same class, along with the loss function of each sub-autoencoder. The equation for the same is given as:

$$\begin{aligned}\underset{\substack{\mathbf{w}_e^1,\ldots,\mathbf{w}_e^k,\\ \mathbf{w}_d^1,\ldots,\mathbf{w}_d^k}}{\arg\min}\ \lambda(\|x_1 - b\circ a(x_1)\|_2^2 + \|x_2 - b\circ a(x_2)\|_2^2)\\ +(1-\lambda)\,\|O_k(x_1) - O_k(x_2)\|_2^2\,.\end{aligned} \tag{4}$$

where, x_1 and x_2 represent two different input samples belonging to the same class. For each sub-autoencoder, $a(x) = \phi(\mathbf{W}_e^k\phi(\mathbf{W}_e^{k-1}...\phi(\mathbf{W}_e^1(x)))$ and $b(x) = \mathbf{W}_d^1(\mathbf{W}_d^2...\mathbf{W}_d^k(x))$, where $\mathbf{W}_e^i$ and $\mathbf{W}_d^i$ refer to the encoding and decoding weights of the i^{th} layer, and $O_k(x)$ is the output of the k^{th} layer.

Recently, Majumdar et al. (2017) present a class sparsity based supervised encoding algorithm wherein a joint-sparsity promoting $l_{2,1}-$norm supervision penalty is added to the loss function. For samples $\mathbf{X}$, belonging to total C classes, the modified algorithm is presented as:

$$\underset{\mathbf{w}_e^1,\ldots,\mathbf{W}_e^k,\mathbf{W}_d^1,\ldots,\mathbf{W}_d^k}{\arg\min}\ \|\mathbf{X} - b\circ a(\mathbf{X})\|_F^2 + \lambda\sum_{c=1}^{C}\|\mathbf{W}_e\mathbf{X}_c\|_{2,1} \tag{5}$$

where, $\mathbf{X}_c$ refers to the samples belonging to class c. The regularization term enforces same sparsity signature across each class, which leads to similar representations of samples from a given class.

2.2. Proposed Class Specific Mean Autoencoder

While all the above techniques incorporate supervision into an otherwise unsupervised model, the proposed architecture incorporates the mean feature of each class into the feature learning process as well. The key motivation behind the proposed algorithm is illustrated in Fig. 4. In this example, with faces as input, *adult* and *minor* as the two classes, the mean adult image and mean minor image are computed using the training samples. For a given probe face image, computing the l_2 distance with respect to the mean minor image provides the similarity of the sample with the minor class. It can be observed that the difference between a *minor* probe image and the mean *minor* image is lower as compared to the difference between an *adult* probe image and the mean *minor* image. This example shows that if the intra-class variations are encoded, it may help in learning class-specific features. Inspired from this observation, in this research, we present a novel formulation of Class Specific Mean Autoencoder.

In the proposed formulation, the loss function of an autoencoder (Hinton and Salakhutdinov, 2006) is updated by introducing class specific information. For simplicity and clarity, Eq. 1

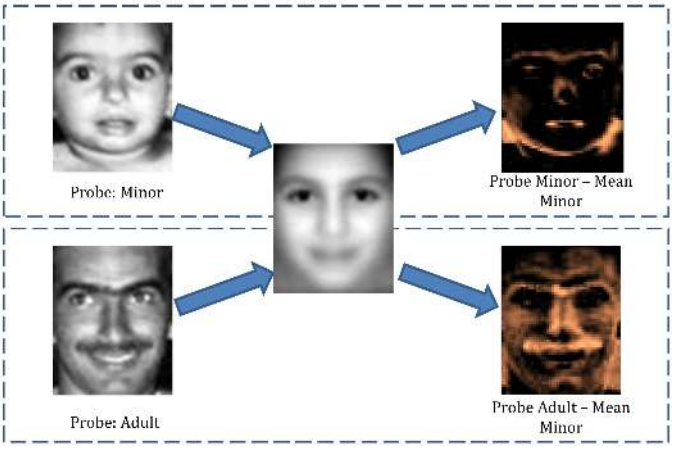


Fig. 4. For a two class problem consisting of *adults* and *minors*, and two probe images, the figure depicts the difference of each probe with respect to the mean minor image. It can be observed that the difference of the minor probe image from the mean minor image is significantly less than the difference of the adult probe image with the mean minor image. This motivates the use of class specific mean feature vectors for incorporating supervision in the feature learning process.

is repeated as follows:

$$\arg\min_{\mathbf{W}_e, \mathbf{W}_d} \|x - \mathbf{W}_d \phi(\mathbf{W}_e x)\|_2^2 \tag{6}$$

For an input sample x_c, belonging to class c, the feature vector f_{x_c} is defined as follows:

$$f_{x_c} = \phi(\mathbf{W}_e x_c) \tag{7}$$

The mean feature vector pertaining to the c^{th} class is defined as:

$$m_c = \mu(\phi(\mathbf{W}_e \mathbf{X}_c)) \tag{8}$$

where, μ represents the mean operator, and $\mathbf{X}_c$ represents all the training samples belonging to class c.

As discussed earlier in this section, we postulate that encoding the difference between the feature of a sample and the mean sample of the same class can help in encoding class-specific features. In other words, the feature of a particular class is brought similar/closer to the *mean* feature of that class. To encode this information, Eqs. 7 and 8 are utilized to form the following optimization constraint:

$$\left\| f_{x_c} - m_c \right\|_2^2 \tag{9}$$

The above equation is incorporated into an autoencoder to create Class Specific Mean Autoencoder as follows:

$$\arg\min_{\mathbf{W}_e, \mathbf{W}_d} \|x_c - \mathbf{W}_d \phi(\mathbf{W}_e x_c)\|_2^2 + \lambda \left\| f_{x_c} - m_c \right\|_2^2 \tag{10}$$

where, λ is the regularization constant. The proposed Class Specific Mean Autoencoder learns the weight parameters such that the features of a particular class are *grouped* together. Expanding Eq. 10, we obtain:

$$\begin{aligned} \arg\min_{\mathbf{W}_e, \mathbf{W}_d} & \|x_c - \mathbf{W}_d \phi(\mathbf{W}_e x_c)\|_2^2 + \\ & \lambda \|\phi(\mathbf{W}_e x_c) - \mu(\phi(\mathbf{W}_e \mathbf{X}_c))\|_2^2 \end{aligned} \tag{11}$$

The updated loss function of Eq. 11 ensures that the learned feature for a sample is close to the mean representation of its class, while being representative of the input sample as well. The second term is added for supervised regularization and can be viewed as:

$$E = \left\| f_{x_c} - t \right\|_2^2 \tag{12}$$

for a given expected target t and obtained output f_{x_c}. The above equation draws a direct parallel with Eq. 1, where the expected target is x, and the obtained output is $(\mathbf{W}_d \phi(\mathbf{W}_e x))$. Similar to the update rule for Eq. 1, the update rule for the above regularization term for j^{th} expected (t_j) and obtained (o_j) output, with respect to weight $w_{e_{i,j}}$, can be written as:

$$\frac{\partial \mathbf{E}}{\partial \mathbf{W}_{e_{i,j}}} = \frac{1}{2} * (o_j - t_j) * \frac{\partial o_j}{\partial \mathbf{W}_{e_{i,j}}} \tag{13}$$

Similar to the gradient descent backpropogation applied to Eq. 1, the Class Specific Mean Autoencoder is solved iteratively via the above update rule till convergence.

For a k layered Class Specific Mean Autoencoder, having encoding weights as $\mathbf{W}_e^1, \mathbf{W}_e^2, ..., \mathbf{W}_e^k$, and decoding weights as $\mathbf{W}_d^1, \mathbf{W}_d^2, ..., \mathbf{W}_d^k$, the loss function of Eq. 10 can be modified as:

$$\arg\min_{\substack{\mathbf{w}_e^1, ..., \mathbf{w}_e^k, \\ \mathbf{w}_{d'}^1, ..., \mathbf{w}_d^k}} \|x_c - b \circ a(x_c)\|_2^2 + \sum_{i=1}^{i=k} \lambda_i \left\| f_{x_c}^i - m_c^i \right\|_2^2 \tag{14}$$

where, $a(x) = \phi(\mathbf{W}_e^k(\phi(\mathbf{W}_e^{k-1} \dots (\phi(\mathbf{W}_e^1 x)))))$ is the encoding function, and $b(x) = \mathbf{W}_d^1(\mathbf{W}_d^2 \dots (\mathbf{W}_d^k x))$ corresponds to the decoding function, and $f_{x_c}^k$ and m_c^k are defined as:

$$f_{x_c}^k = \phi(\mathbf{W}_e^k(\phi(\mathbf{W}_e^{k-1} \dots (\phi(\mathbf{W}_e^1 x_c))))) \tag{15}$$

$$m_c^k = \mu(\phi(\mathbf{W}_e^k(\phi(\mathbf{W}_e^{k-1} \dots (\phi(\mathbf{W}_e^1 \mathbf{X}_c)))))) \tag{16}$$

Owing to the large number of parameters involved, the optimization of the above model is performed via the greedy layer by layer approach (Bengio et al., 2007). At the time of testing, the learned encoding weights $(\mathbf{W}_e^1, \mathbf{W}_e^2, \dots \mathbf{W}_e^k)$ are used to calculate the feature vector for a given sample, which is then provided as input to a classifier. Fig. 5 presents a pictorial representation of the proposed algorithm, for a two class problem.

2.3. Predicting Adulthood using Class Specific Mean Autoencoder

The proposed Class Specific Mean Autoencoder is used to address the problem of classification of face images into adults (18 years of age or more) or minors (less than 18 years of age). The proposed model is used for feature extraction, which is then followed by a Neural Network for classification. The algorithm is summarized in Algorithm 1.

3. Datasets and Experimental Details

Given a face image, predicting whether the individual is an adult or not, can be modeled as a two class classification problem: individuals below the age of 18 years are referred as *minors*, while individuals of age equal to or greater than 18 years are referred as *adults*. Details regarding the datasets used, and the experimental protocols are as follows:

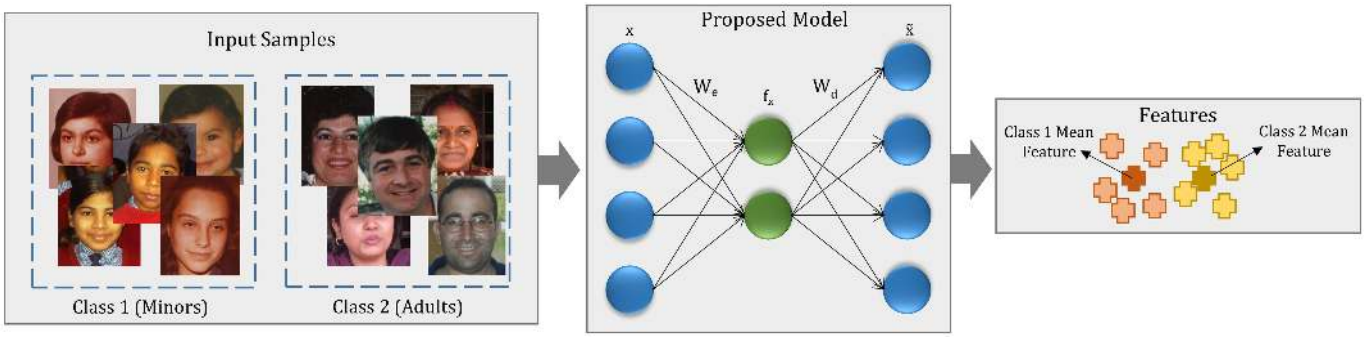


Fig. 5. Proposed Class Specific Mean Autoencoder. x **and** $\tilde{x}$ **represent the input and the reconstructed sample respectively,** $\mathbf{W}_e$ **and** $\mathbf{W}_d$ **denote the encoding and decoding weights, and** f_x **corresponds to the learned feature vector.**

Algorithm 1: Training Single Layer Class Specific Mean Autoencoder for Adulthood Prediction

Input : Training images of minor ($\mathbf{X}_{minor}$) and adult ($\mathbf{X}_{adult}$) classes, iter = 0, maxIter.
Output: Encoding and decoding weights: $\mathbf{W}_e$, $\mathbf{W}_d$.

Initialize $\mathbf{W}_e$ and $\mathbf{W}_d$;
while *iter* < *maxIter* **do**
  Compute mean adult feature (m^{iter}_{adult}) using Eq. 8 ;
  Compute mean minor feature (m^{iter}_{minor}) using Eq. 8 ;
  foreach $x_{minor} \in \mathbf{X}_{minor}$ **do**
    Minimize Eq. 10 using x_{minor} and m^{iter}_{minor};
  end
  foreach $x_{adult} \in \mathbf{X}_{adult}$ **do**
    Minimize Eq. 10 using x_{adult} and m^{iter}_{adult};
  end
  iter++;
end

(a) Multi-Ethnicity Dataset

(b) MORPH Album-II Dataset

Fig. 6. Sample images from the datasets used for experimental evaluation.

3.1. Datasets Used

Experiments are performed on two datasets: (i) Multi-Ethnicity dataset and (ii) MORPH Album-II dataset. Fig. 6 shows some sample images from both the datasets. Details about each are given below:

3.1.1. Multi-Ethnicity Dataset

Since the existing datasets containing images of minors and adults contain very limited variations with respect to ethnicity, pose, and expression, along with very few samples below the age of 16, we propose the Multi-Ethnicity Dataset. Multi-Ethnicity dataset consists 13,133 face images combined from

- Proposed Multi-Resolution Face Dataset containing 4,019 face images,
- Heterogeneous Dataset containing 8,112 face images (Dhamecha et al., 2011), and
- FG-Net Aging Dataset containing 1,002 face images (Panis et al., 2016).

The dataset contains variations across ethnicity, gender, resolution, illumination, as well as minute pose and expression. Due to the lack of datasets containing face images of both adults and minors, we have also created *Multi-Resolution Face Dataset (MRFD)* consisting of 4,019 face images of minors and adults of two resolutions with slight variations in pose, expression, and illumination. The Multi-Resolution Face Dataset[1] consists of 4,019 Indian face images of 317 subjects captured in outdoor, as well as indoor environment. The dataset consists of images of 307 minors (3,896 images) and 10 adults. Images have been captured from two smartphones (with 3.1 MP camera) with resulting face size of 360×420 pixels, and a high resolution hand-held Canon digital camera with resulting face images of dimension 560×680. Each subject has at least 12 near-frontal, well-illuminated images (at least 4 from each camera source). The dataset contains variations in age, ranging from toddlers to adults of around 50 years. The subjects were only asked to look at the camera, without any instructions for pose or expression, which resulted in images with varying head movement and expression. This is the first dataset containing such large number of minor images which would help in facilitating research on minor face images as well. Fig. 7 presents some sample images from the MRF dataset illustrating variations in age, illumination, and expression.

3.1.2. MORPH Album-II Dataset

Craniofacial Longitudinal Morphological Face (MORPH) dataset (Ricanek Jr. and Tesafaye, 2006) consists of two al-

[1]The dataset will be made publicly available for academic research via http://iab-rubric.org/resources.html

Table 2. Summarizing the dataset description and experimental protocol.

Database	Total Number of Images	Number of Images of		Number of Images in	
		Minors	Adults	Train Set	Test Set
MORPH Album-II Dataset	55,132	3,330	51,802	4,662	50,470
Multi-Ethnicity Dataset	13,133	8,574	4,559	6,276	6,857

Fig. 7. Sample images from the proposed Multi-Resolution Face dataset.

bums: Album-I contains scanned digital face images, while Album-II contains longitudinal digital face images captured over several years. A subset of Album-II containing 55,134 images of 13,000 subjects is made available for academic researchers, which has been used for experimental analysis in this research. The dataset contains images of subjects between the age range of 16 to 77 years, and also provides metadata for race, gender, date of birth, and date of acquisition.

3.2. Experimental Protocol

Unseen training and testing partitions are created for both the datasets. For training, equal number of samples from both the classes are used, which is defined by the class with lesser number of samples. 70% of the samples corresponding to the minor class and equal number of images from adult class are used for training, while the remaining data is used for testing. For the MORPH dataset, this results in the training and testing sets of size 4,662 and 50,470 images respectively. Similarly, for the Multi-Ethnicity dataset, 6,276 images are selected for training with the constraint that equal number of samples are selected from both the classes. The remaining face images constitute the test set. Details of data partitioning are documented in Table 2.

To showcase the efficacy of the proposed algorithm, comparison has been drawn with other deep learning based feature extractors; namely, Stacked Denoising Autoencoder (SDAE) (Vincent et al., 2010), Deep Boltzmann Machine (DBM) (Hinton, 2012), and Discriminative Restricted Boltzmann Machine (DRBM) (Larochelle and Bengio, 2008). Comparison has also been drawn with VGG-Face descriptor (Parkhi et al., 2015), which is one of the state-of-the-art deep learning based feature extractor. Features extracted from these models are provided as input to a Neural Network for classification. A CNN based Commercial-Off-The-Shelf (COTS) system, Face++ (Zhou et al., 2015), has also been used to compare the performance of the proposed model. Since there does not exist any COTS for the task of adulthood prediction, Face++ is used to estimate the age of the given face image, which is then utilized to classify the input as an adult or a minor. In order to analyze the statistical significance of the results obtained by the proposed model, McNemar test (McNemar, 1947) has been performed. Given the classification results obtained from two models, McNemar test predicts whether the performance of both the models is statistically different or not. For every comparison of the proposed Class Specific Mean Autoencoder with an existing architecture, a p-value is reported. A smaller p-value corresponds to a higher confidence level of statistical difference. In this research, all claims of statistical significance have been made at a confidence level of 95%.

3.3. Implementation Details

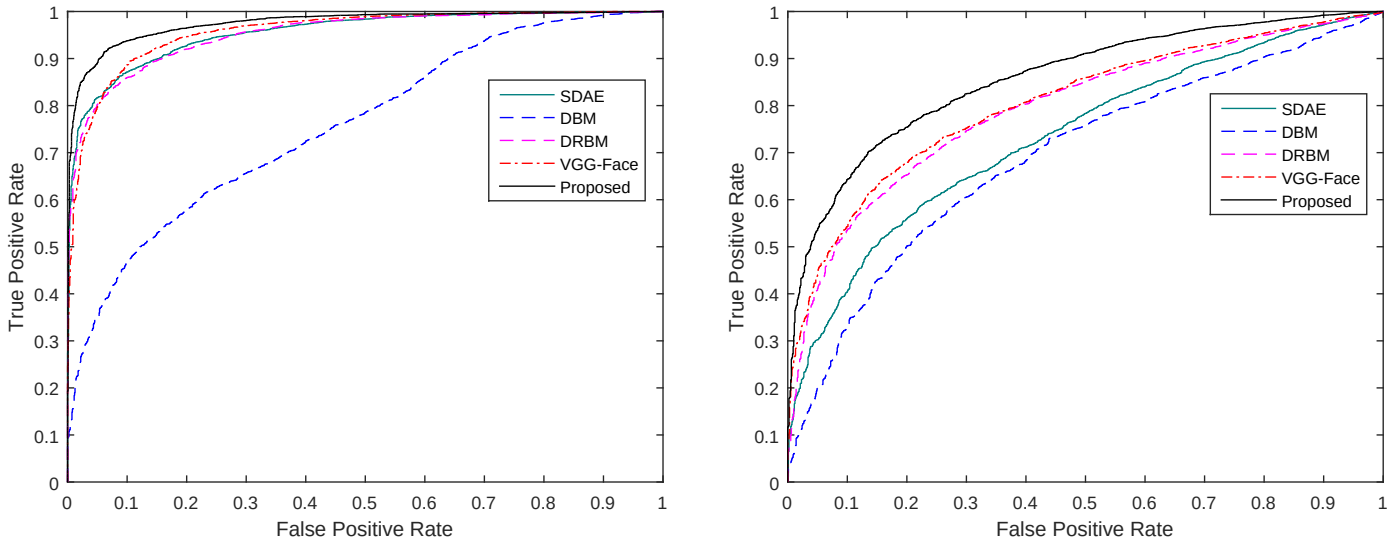
For all the experiments, face detection is performed on all images using Viola Jones Face Detector (Viola and Jones, 2004), following which the images are geometrically normalized and resized to a fixed size. A Class Specific Mean Autoencoder of dimensions $[m, m]$ is learned, where m is the size of the image. Following this, a neural network of dimension $[\frac{m}{4}, \frac{m}{8}]$ is trained for classification. *sigmoid* activation function is used at the hidden layers. Models are trained for 100 epochs with a learning rate of 0.01. We have followed the best practices used for setting the parameters and architecture for deep learning (Larochelle et al., 2009). For existing algorithms, in order to maintain consistency, a two layer architecture is utilized for the feature extractor and neural network.

4. Experimental Results and Observations

Owing to the large class imbalance in the test samples, mean class-wise accuracy has been reported for all the experiments. The formula used for calculating the accuracy is as follows:

$$Accuracy = \frac{Accuracy_{Minor} + Accuracy_{Adult}}{2} \tag{17}$$

where, $Accuracy_{Minor}$ and $Accuracy_{Adult}$ correspond to the accuracies obtained for minor and adult classification, respectively by a particular model.



(a) Multi-Ethnicity Dataset

(b) MORPH Album-II Dataset

Fig. 8. Receiver Operating Characteristic (ROC) curves obtained for categorizing whether a given face image is of an adult or not.

4.1. Results on Multi-Ethnicity Dataset

Table 3 presents the classification accuracies of the proposed model along with other existing architectures on the Multi-Ethnicity Dataset. Fig. 8 also presents the Receiver Operative Characteristic (ROC) curves obtained for the experiments. It is observed that the proposed Class Specific Mean Autoencoder (2-layer) achieves a classification accuracy of **92.09%**, which is at least 2.5% better than existing algorithms. This is followed by VGG-Face with an accuracy of 89.45%, while Face++ (commercial off-the-shelf system) achieves a classification performance of 78.41%. The improvement of 5.29% in performance of the proposed model as compared to a Stacked Denoising Autoencoder can be attributed to the additional class representative terms added to the autoencoder formulation.

Table 3 also presents the p-values obtained upon performing the McNemar test to evaluate the statistical difference. Since all the values are below 0.05, we can claim with a confidence level of 95% that the performance of the proposed model is statistically different from all other existing models. In order to understand the effect of number of layers, experiments are also performed using a single layer Class Specific Mean Autoencoder. For a single layer, the proposed model yields an accuracy of 91.58%, which continues to show an improvement of at least 2%, compared to other models with the same architecture.

To analyze the class-specific classification accuracies, Table 5 presents the confusion matrix for the proposed Class Specific Mean Autoencoder on the Multi-Ethnicity dataset. The results indicate that the performance of the trained model is not biased towards any particular class by achieving a classification accuracy of 93.52% and 90.65% on the two classes of adults and minors, respectively. This is essential to ensure that while unauthorized access is not provided to minors, rightful adults are not restricted from it either. In order to cater to the application of age-specific authorized access control, it is essential to ensure that the percentage of people below the age of majority, i.e. minor, obtaining unauthorized access should be minimal. To analyze the performance of all architectures for such an application, Fig. 11 presents bar graphs summarizing the percentage of minors misclassified as adults. It can be seen that the proposed model achieves a misclassification percentage of 9.35%, as opposed to 22.97% by Face++. Fig. 9 presents some sample images from the Multi-Ethnicity dataset misclassified as adults by *all* the algorithms. It can be observed from the sample images that these images were captured either near the age of majority of 16-17 years or have artifacts such as headbands/scarves, which further make the task of adulthood classification challenging. Certain samples of kids below the age of one year were also mis-classified, possibly due to the undeveloped features of newborns.

Fig. 9. Sample images from the Multi-Ethnicity dataset, incorrectly classified by all algorithms. At the time of capture, all individuals were below the age of 18. It can be seen that while the actual age of the samples was below the age of majority, it is easy to mistake minors of 16-17 years as adults. External accessories such as scarves may also introduce mis-classification, resulting in unauthorized access control.

The major challenges associated with the problem of adult classification lie in the age bracket of 16 to 19 years (16-17: minors, and 18-19: adults). On the Multi-Ethnicity dataset, the proposed algorithm achieves a classification accuracy of 64.58% on the above mentioned age range. VGG-Face, which performs the second best, reports an accuracy of 58.33%, which is at least 6% lower than the proposed algorithm. Fig. 10 displays sample images from the specified age range and are mis-

Table 3. Classification Accuracy (%) on Multi-Ethnicity dataset. p-Value corresponds to the values obtained after performing McNemar test to compare the classification performance of an existing architecture with the proposed Class Specific Mean Autoencoder. The proposed model presents improved classification performance, while being statistically different from all other models at a confidence level of 95%.

Method	Accuracy (%)	p-Value	Statistical Significance
SDAE (Vincent et al., 2010)	86.80	0.003	Significant
DBM (Hinton, 2012)	65.16	< 0.001	Significant
DRBM (Larochelle and Bengio, 2008)	87.03	0.001	Significant
VGG-Face (Parkhi et al., 2015)	89.45	0.004	Significant
COTS: Face++ (Zhou et al., 2015)	78.41	< 0.001	Significant
Proposed Class Specific Mean Autoencoder	**92.09**	-	-

Table 4. Classification Accuracy (%) on MORPH Album-II dataset. p-Value corresponds to the values obtained after performing McNemar test to compare the classification performance of an existing architecture with the proposed Class Specific Mean Autoencoder. The proposed model presents improved classification performance, while being statistically different from all other models at a confidence level of 95%.

Method	Accuracy (%)	p-Value	Statistical Significance
SDAE (Vincent et al., 2010)	66.25	0.005	Significant
DBM (Hinton, 2012)	65.30	< 0.001	Significant
DRBM (Larochelle and Bengio, 2008)	65.72	< 0.001	Significant
VGG-Face (Parkhi et al., 2015)	70.44	0.010	Significant
COTS: Face++ (Zhou et al., 2015)	57.23	< 0.001	Significant
Proposed Class Specific Mean Autoencoder	**73.13**	-	-

Fig. 10. Sample images from the Multi-Ethnicity dataset that are in the age bracket of 16-19 years and misclassified by the proposed Class Specific Mean Autoencoder. The first image belongs to an adult of age 19 years, while the remaining belong to entities below the age of majority.

Table 5. Confusion matrix of Class Specific Mean Autoencoder on the Multi-Ethnicity database.

		Predicted	
		Adult	Not Adult
Actual	Adult	93.52%	6.48%
	Not Adult	9.35%	90.65%

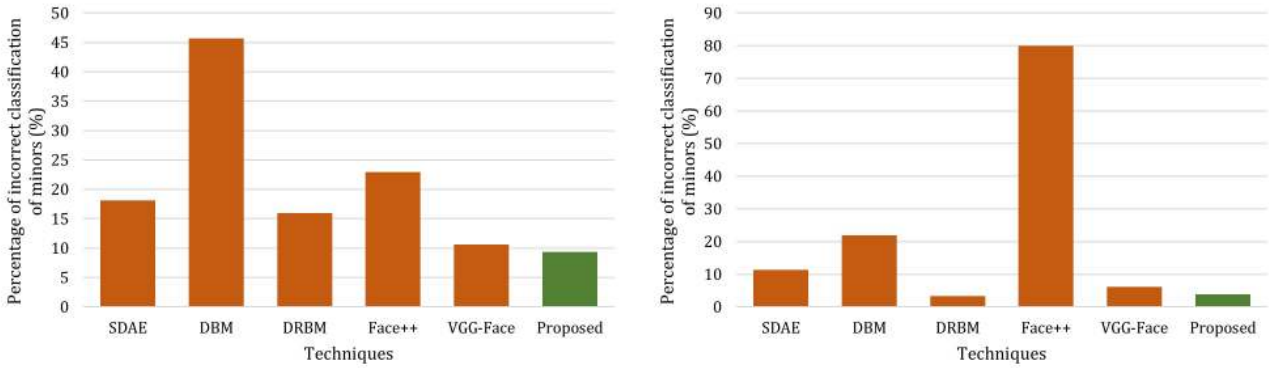
classified by the proposed algorithm. The images demonstrate the challenging nature of human aging which are dependent on intrinsic and extrinsic person-specific factors, such as health, environment, and climate.

4.2. Results on MORPH Album-II Dataset

The classification accuracies obtained by the proposed model and other existing architectures are tabulated in Table 4, and Fig. 8 presents the Receiver Operating Characteristic (ROC) curves obtained for the experiments. It is observed that the proposed architecture achieves a classification accuracy of **73.13%**, which is at least 2.5% better than existing approaches, while Face++ (COTS) achieves an accuracy of 57.23%. Table 4 also presents the p-values obtained upon performing the McNemar statistical test on the proposed Class Specific Mean Autoencoder and other existing models. While the second best performance is achieved by VGG-Face features (70.44%), it is important to note that the improvement in accuracy achieved by the proposed model is statistically significant for a confidence level of 95%. Upon analyzing the gender-specific adulthood prediction results, it can be observed that the classification accuracy on female sample images is 62.89%, whereas the accuracy on male sample images is 75.09%. It is further observed that for females, the misclassification of adults as minors is much higher, as compared to males, thereby resulting in an overall lower classification performance.

From Fig. 11, it can be observed that the proposed model achieves a *minor* misclassification percentage of only 3.9%, as opposed to nearly 80% by Face++ (COTS) on the MORPH Album-II dataset. The high misclassification rate of minors by Face++ reinstates the requirement for robust algorithms with the ability to process and analyze minor face images as well. It is important to note that the age of face images in MORPH Album-II dataset varies from 16 to 77 years. Thus, resulting in a further challenging dataset having multiple subjects *just below* the age of majority. The higher misclassification rate achieved can thus also be attributed to this challenging age range. It is also interesting to observe from Fig. 11 that while DRBM achieves a lower misclassification of minors as adults (i.e. 3.30%), the overall classification accuracy of DRBM based approach is less than the proposed approach (Table 4). This further motivates the use of the proposed algorithm for ensuring rightful access to adults, while restricting minors.

Fig. 12 presents sample images of the age group of 16-19 (16-17: minors, 18-19: adults) years from the MORPH Album-II dataset which are correctly identified by the proposed algorithm and not by any other algorithm. Upon analyzing the mean images of both the classes, we observe a significant visual difference in the jaw area of minors and adults. As can be seen from Fig. 12 as well, minors appear to have a tighter jaw line which is often not observed with adults. We believe that this variation has been encoded well by the proposed model among



(a) Multi-Ethnicity Dataset

(b) MORPH Album-II Dataset

Fig. 11. Percentage of minors incorrectly classified as adults for both the datasets. A lower percentage would ensure fewer instances of unauthorized access.

Table 6. Classification accuracy (%) on perturbed face images for Multi-Ethnicity and MORPH Album-II datasets.

Perturbation	Multi-Ethnicity Dataset		MORPH Album-II Dataset	
	Proposed	VGG-Face	Proposed	VGG-Face
No Perturbation (Original)	92.09	89.45	73.13	70.44
Gaussian Blur (Sigma = 3)	89.54	61.34	72.40	50.00
Gaussian Noise (Mean = 0, Std. dev. = 0.01)	87.95	50.59	72.09	50.03
Gaussian Noise(Mean = 0, Std. dev. = 0.001)	91.66	60.57	72.98	62.08
Holes (10 holes of 3×3)	87.35	64.67	72.55	56.70

(a) Minors

(b) Adults

Fig. 12. Sample images from the MORPH Album-II dataset correctly classified by the proposed algorithm, and not by other existing algorithms. (a) shows images of individuals having age 16 (first two samples) or 17, whereas (b) depicts just turned adults of 18 (first two) or 19 years of age.

other features, resulting in superior performance.

4.3. Performance on Perturbed Face Images

It has often been observed in literature that the performance of deep models deteriorates in the presence of perturbations (Goswami et al., 2018). The proposed model has also been evaluated on perturbed face images in order to understand its vulnerabilities. This is performed by incorporating perturbations in the form of Gaussian blur, Gaussian noise, and holes in the original face images. Experiments are performed on the Multi-Ethnicity and the MORPH Album-II datasets with the protocols discussed earlier. The models are trained on unperturbed (original) images but the test images are perturbed. In this evaluation, no separate training is performed for the perturbed face images. Table 6 presents the classification accuracies obtained from the proposed Class Specific Mean Autoencoder, and the second best performing model, VGG-Face. It can be observed that with perturbed test images, the accuracy of the proposed model reduces by less than 5% and 1.04% for Multi-Ethnicity and MORPH Album-II datasets, respectively. On the other hand, VGG-Face demonstrates a drop of at least 24% and 8% on the two datasets, respectively. This experiment demonstrates the utility of the proposed model for performing classification under different kinds of perturbations.

5. Conclusion

Faces are often seen as a viable non-invasive modality for predicting the age of an individual. However, due to the large intra-class variations, predicting adulthood from face images is an arduous task. The key contribution of this research is developing a novel formulation for *Class Specific Mean Autoencoder* and utilize it for adulthood classification. The proposed formulation aims to learn supervised feature vectors that maximize the intra-class similarity. Experimental results and comparison with existing approaches on two large databases: the proposed Multi-Ethnicity dataset and MORPH Album-II dataset showcase the effectiveness of the proposed algorithm. In future, we plan to extend the proposed formulation to incorporate multiclass-multilabel information in feature learning.

References

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2007. Greedy layer-wise training of deep networks, in: Advances in Neural Information Processing Systems 19. MIT Press, pp. 153–160.

Chen, D., Ren, S., Wei, Y., Cao, X., Sun, J., 2014. Joint cascade face detection and alignment, in: European Conference on Computer Vision, pp. 109–122.

Dhamecha, T.I., Sankaran, A., Singh, R., Vatsa, M., 2011. Is gender classification across ethnicity feasible using discriminant functions?, in: IEEE International Joint Conference on Biometrics, pp. 1–7.

Eidinger, E., Enbar, R., Hassner, T., 2014. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security vol. 9, no. 12, 2170–2179.

Escalera, S., Fabian, J., Pardo, P., Baró, X., Gonzàlez, J., Escalante, H.J., Misevic, D., Steiner, U., Guyon, I., 2015. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results, in: IEEE International Conference on Computer Vision Workshop, pp. 243–251.

Farfade, S.S., Saberian, M.J., Li, L.J., 2015. Multi-view face detection using deep convolutional neural networks, in: International Conference on Multimedia Retrieval, pp. 643–650.

Ferguson, E., Wilkinson, C., 2017. Juvenile age estimation from facial images. Science & Justice 57, 58 – 62.

Fu, Y., Guo, G., Huang, T.S., 2010. Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 32, no. 11, 1955–1976.

Gao, S., Zhang, Y., Jia, K., Lu, J., Zhang, Y., 2015. Single sample face recognition via learning deep supervised autoencoders. IEEE Transactions on Information Forensics and Security vol. 10, no. 10, 2108–2118.

Geng, X., Yin, C., Zhou, Z.H., 2013. Facial age estimation by learning from label distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 35, no. 10, 2401–2412.

Goswami, G., Ratha, N., Agarwal, A., Singh, R., Vatsa, M., 2018. Unravelling robustness of deep learning based face recognition against adversarial attacks, in: AAAI.

Guo, G., Fu, Y., Dyer, C.R., Huang, T.S., 2008. Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Transactions on Image Processing vol. 17, no. 7, 1178–1188.

Han, H., Otto, C., Jain, A.K., 2013. Age estimation from face images: Human vs. machine performance, in: International Conference on Biometrics, pp. 1–8.

Hinton, G., Salakhutdinov, R., 2006. Reducing the dimensionality of data with neural networks. Science vol.313, no.5786, 504 – 507.

Hinton, G.E., 2012. A practical guide to training restricted boltzmann machines, in: Neural Networks: Tricks of the Trade - Second Edition, pp. 599–619.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics vol.22, no.1, 79–86.

Larochelle, H., Bengio, Y., 2008. Classification using discriminative restricted boltzmann machines, in: International Conference on Machine Learning, pp. 536–543.

Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P., 2009. Exploring strategies for training deep neural networks. Journal of Machine Learning Research vol. 10, 1–40.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Levi, G., Hassner, T., 2015. Age and gender classification using convolutional neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 34–42.

Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G., 2015. A convolutional neural network cascade for face detection, in: IEEE Conference on Computer Vision and Pattern Recognition.

Li, K., Xing, J., Hu, W., Maybank, S.J., 2017. D2c: Deep cumulatively and comparatively learning for human age estimation. Pattern Recognition 66, 95 – 105.

Luu, K., Ricanek, K., Bui, T.D., Suen, C.Y., 2009. Age estimation using active appearance models and support vector machine regression, in: IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–5.

Majumdar, A., Singh, R., Vatsa, M., 2017. Face verification via class sparsity based supervised encoding. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1273–1280.

Martinez, J.A., Rutledge, P.C., Sher, K.J., 2007. Fake id ownership and heavy drinking in underage college students: Prospective findings. Psychology of addictive behaviors 21, 226.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika vol.12, no.2, 153–157.

Meng, Q., Catchpoole, D., Skillicom, D., Kennedy, P.J., 2017. Relational autoencoder for feature extraction, in: International Joint Conference on Neural Networks, pp. 364–371.

Ng, A., 2011. Sparse autoencoder. CS294A Lecture notes 72, 1–19.

Panis, G., Lanitis, A., Tsapatsoulis, N., Cootes, T.F., 2016. Overview of research on facial ageing using the FG-NET ageing database. IET Biometrics 5, 37–46.

Parkhi, O.M., Vedaldi, A., Zisserman, A., 2015. Deep face recognition, in: British Machine Vision Conference.

Pontes, J.K., Britto, A.S., Fookes, C., Koerich, A.L., 2016. A flexible hierarchical approach for facial age estimation based on multiple features. Pattern Recognition 54, 34 – 51.

Ricanek Jr., K., Tesafaye, T., 2006. MORPH: A longitudinal image database of normal adult age-progression, in: International Conference on Automatic Face and Gesture Recognition, pp. 341–345.

Rifai, S., Mesnil, G., Vincent, P., Muller, X., Bengio, Y., Dauphin, Y., Glorot, X., 2011a. Higher order contractive auto-encoder, in: European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 645–660.

Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y., 2011b. Contractive auto-encoders: Explicit invariance during feature extraction, in: International Conference on Machine Learning, pp. 833–840.

Rothe, R., Timofte, R., Gool, L.V., 2015. DEX: Deep expectation of apparent age from a single image, in: IEEE International Conference on Computer Vision Workshop, pp. 252–257.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Li, F., 2014. Imagenet large scale visual recognition challenge. CoRR abs/1409.0575. URL: `http://arxiv.org/abs/1409.0575`.

Schroff, F., Kalenichenko, D., Philbin, J., 2015. FaceNet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. CoRR abs/1409.1556. URL: `http://arxiv.org/abs/1409.1556`.

Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. Deepface: Closing the gap to human-level performance in face verification, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research vol. 11, 3371–3408.

Viola, P., Jones, M.J., 2004. Robust real-time face detection. International Journal of Computer Vision vol. 57, no. 2, 137–154.

Wang, W., Huang, Y., Wang, Y., Wang, L., 2014. Generalized autoencoder: A neural network framework for dimensionality reduction, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 496–503.

Wechsler, H., Jae, E., Nelson, T., Kuo, M., 2002. Underage college students' drinking behavior, access to alcohol, and the influence of deterrence policies: Findings from the harvard school of public health college alcohol study. Journal of American College Health vol.50, no.5, 223–236.

Xing, J., Li, K., Hu, W., Yuan, C., Ling, H., 2017. Diagnosing deep learning models for high accuracy age estimation from a single image. Pattern Recognition 66, 106 – 116.

Zhang, X., Fu, Y., Jiang, S., Sigal, L., Agam, G., 2015. Learning from synthetic data using a stacked multichannel autoencoder, in: International Conference on Machine Learning and Applications, pp. 461–464.

Zheng, X., Wu, Z., Meng, H., Cai, L., 2014. Contrastive auto-encoder for phoneme recognition, in: International Conference on Acoustics, Speech and Signal Processing, pp. 2529–2533.

Zhou, E., Cao, Z., Yin, Q., 2015. Naive-deep face recognition: Touching the limit of LFW benchmark or not? CoRR abs/1501.04690. URL: `http://arxiv.org/abs/1501.04690`.

Zhuang, F., Cheng, X., Luo, P., Pan, S.J., He, Q., 2015. Supervised representation learning: Transfer learning with deep autoencoders, in: International Joint Conference on Artificial Intelligence, pp. 4119–4125.