On Automatic Age Estimation from Facial Profile View

Ali Maina Bukar [1*] and Hassan Ugail[1]
[1] Center for Visual Computing, University of Bradford, Richmond Road, Bradford, United Kingdom.
[*]ambukar@student.bradford.ac.uk

Abstract: In recent years, automatic facial age estimation has gained popularity due to its numerous applications. Much work has been done on frontal images and lately, minimal estimation errors have been achieved on most of the benchmark databases. However, in reality, images obtained in unconstrained environments are not always frontal. For instance, when conducting a demographic study or crowd analysis, one may get profile images of the face. To the best of our knowledge, no attempt has been made to estimate ages from the side-view of face images. Here we exploit this by using a pre-trained deep residual neural network (ResNet) to extract features. We then utilize a sparse partial least squares regression approach to estimate ages. Despite having less information as compared to frontal images, our results show that the extracted deep features achieve a promising performance.

1. Introduction

The human face carries huge amount of information. It has been well documented that facial structures and appearances change considerably as people age [1], making it difficult to recognize individuals over long periods of time. In particular, the shape of the face changes substantially from birth to adulthood. During adulthood, while the shape remains relatively constant, there are changes in musculature and skin tautness affecting the facial 'texture' [2]. Thus, it is no surprise that carefully trained computer algorithms have been able to estimate human ages with minimal errors using frontal face images [3].

The basic goal of an age estimation algorithm is to map an image of a person's face into an estimate of their age. The classical approach to estimation follows a two-step procedure; feature extraction and age-pattern learning. Over the last two decades, various feature extraction techniques have been proposed by researchers. These include active appearance model (AAM) [4], histograms of gradients (HOG) [5], LBP operators [6], Gabor filters [7], and the Bio-inspired features [8]. Thereafter, age-pattern learning is treated as either a classification or regression problem [9]. Recently, convolutional neural networks (ConvNets) have been used to replace hand crafted feature extraction techniques. Thus, different deep learning architectures [3], [10], [11] have been deployed for end to end learning.

Research efforts have been directed towards frontal or near frontal images even though these images can be difficult to obtain in real life scenarios, such as crowd analysis. To this end, we propose to investigate age estimation from profile images, i.e. face photographs where only one eye and half of the lip are visible. Our approach uses the weights of a pre-trained deep neural network for feature extraction, and sparse partial least squares regression for age estimation.

Our contributions are summarized as follows:

1. A novel attempt to achieve age estimation from profile face images is proposed.
2. Due to limited availability of labelled profile images, we deploy off-the-shelf deep residual neural network features for face representation.
3. Extensive evaluation of the proposed method is conducted on controlled, semi-controlled and unconstrained images.
4. Despite the absence of similar work, the results of our experiments are compared to other state-of-the-art algorithms.

2. Related Work

Age estimation has been extensively studied due to its numerous practical applications, which include demographic studies, crowd analysis, forensic, gaming, and biometrics. Work done on age estimation tends to follow a two-stage process: feature extraction and pattern learning. The former entails techniques used to parameterise the human face, while the latter involves machine learning methods used to classify/estimate ages. In recent times, convolutional neural networks (CNNs/ConvNets) have made it possible to combine these two steps into a single procedure. Thus, features are learnt automatically from raw pixels; thereafter age estimation is conducted using simple classifiers.

In this section, succinct reviews of age estimation and deep neural network features are presented.

2.1. Age Estimation

The earliest work on age estimation can be traced back to Kwon & Lobo [12].They categorised images into babies, young adults and senior adults, through the use of facial anthropometric features. Thereafter, various researchers extended the idea of using facial geometric features [13]–[15]. However, people have since realised that anthropometric features only encode the face shape [2]. One of the earliest works that took into account both shape and texture information of the face is that of Lanitis *et al.* [16], features extracted via active appearance models (AAM), were regressed with ages using a quadratic model. Subsequently, several works have deployed AAM features [4], [17], [18]. Subspace learning techniques have also been utilised in the literature, these include PCA, neighbourhood preserving projections (NPP), locality preserving projection (LPP)

and orthogonal LPP [19]. In the last decade, biologically inspired features (BIF) have been the most widely used feature extraction procedure [5, 6, 20]. Usually, the extracted features were then fed into generic classifiers or regressors. With the recent successes of ConvNets, various deep learning models have also been proposed to tackle the problem of age estimation [10], [11]. Besides the problem of real age estimation, researchers [22]–[24] now study apparent age estimation which is aimed at answering the question of how old does a person look like in a given picture.

A common peculiarity we find in all previous works is that frontal and near frontal images were used for training and testing the algorithms. However, in real life scenarios, frontal images are not always gettable. Thus, contrary to previous works, we propose to investigate age estimation using only profile images. We therefore, aim to answer the following questions; can the computer learn age related information from images that have one-half of the face completely occluded? Is the estimation accuracy comparable to that of frontal-face age estimation?

2.2. Deep Features

The recent success of deep ConvNets has revolutionised computer vision [25] through achieving state-of-the-art results on challenging vision tasks [26], [27]. However, the performance of ConvNets greatly depends on many factors, among which are the training data size and depth of the network [28]. Due to the millions of parameters required for optimisation, training with insufficient data leads to over-fitting. However, research [29], [30] has shown that weights learnt by ConvNets trained on large datasets, can be transferred to other computer vision tasks, especially for scenarios with deficient data. Thus, it is possible to take out certain layers of a trained model (usually the last output layer) and then use the activations of the lower layers, as fixed feature extractors. Using these deep features, several researchers have attained promising outcomes on different applications, without having to train a ConvNet from scratch.

In one of the earliest works, Razavian *et al*. [31] conducted various visual classification tasks using features extracted from the OverFeat model [32]. Interestingly, they achieved superior results even when the target datasets differed from the base dataset that OverFeat was trained to solve. Subsequently, a considerable amount of work has been done using off-the-shelf ConvNet features including, but not limited to, action recognition [33], image retrieval [34] and medical image classification [35]. Astoundingly, in some circumstances, these features outperform specialised ConvNets that are trained from scratch [33], [36].

The depth of a ConvNet has a crucial effect on its performance, research [27] has shown that as the depth increases, accuracy increases until saturation is reached, thereafter performance degradation occurs. In other words, increasing layers via the traditional approach of simply stacking convolution blocks results in higher training and test errors, when the model gets excessively deep. As a result of this problem, deep residual neural network (ResNet) [27] was introduced in 2015. The framework introduces a simple way of increasing network depth, without having to encounter optimization and vanishing gradient problems.

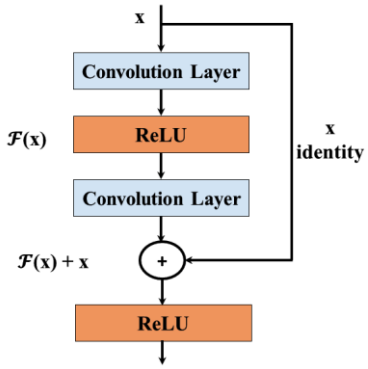


Fig. 1. *Building Block of ResNet*

Developed by Microsoft Research Asia, the deep residual network (ResNet) improves the conventional neural network by introducing an identity shortcut connection in parallel to the normal convolution block (see Fig. 1). Through the added shortcut, back-propagation gets easier, thus making the learning process (convergence) faster. Most importantly, as opposed to plain ConvNets, the addition of layers almost always results in considerable increase in accuracy thereby making it possible to create extremely deep networks [27].

The idea described by [27] is, it is possible to fit a desired underlying mapping $H(x)$ by a few stacked layers, in such a way that they can fit another underlying mapping $F(x) = H(x) - x$. Thus, it is possible to reformulate the mapping $H(x)$ in terms of the input x and the residual function $F(x)$ so that, $H(x) = F(x) + x$. As shown in Fig. 1, the connection of the input to output is called an identity mapping. This reformulation helps to tackle the degradation problem [37] which suggests that solvers have difficulty in approximating identity mappings from multiple nonlinear layers. However, with this new approach, in situations where the optimal function is further away from the zero mapping, as compared to the identity mapping, it will be easier for the solver to find a perturbation with reference to the identity mapping, than learning a new function. Interestingly, the shortcut connection does not introduce additional parameters, nor does it increase computational complexity. More so, it addresses the degradation problem, thereby resulting in extremely deep networks with accuracy gains that are proportionate to their depth.

To show the power of deep residual networks, He *et al.* [27] trained four different architectures (34, 50, 101, and 152 layered ResNets) on the ImageNet [38] dataset. Because of their simple approach to increasing network depth, ResNets eventually won the ImageNet ILSVRC2015, detection, classification, and localization tasks, with a phenomenal error rate of 3.6%.

Inspired by the aforementioned works that used deep features, and by the recorded success of deep residual networks, we hereby propose to use deep features from Microsoft's ResNet model to extract facial features.

3. Method

Our approach is to extract face features, reduce the dimension of the resulting vectors, and then perform regression in order to predict ages.

3.1. Feature Extraction

For face representation, activations of the last average pooling layer of the ResNet model are employed. Precisely, three sets of features are retrieved using ResNet-50, RestNet-101 and ResNet-152 neural networks that were trained on the ImageNet dataset. Contrary to previous works, we have decided not to consider the output of the last fully connected layer, because the dataset the networks were trained on differ from our target data. Moreover, research [33] has shown that lower layers of the deep neural network learn more generic features.

3.2. Choice of Regression Algorithm

Various regression algorithms have been proposed in the literature, relevance vector machine regressor, quadratic model, partial least squares regression (PLS), and support vector regression, to name but a few. PLS regression is one method that has been widely used for regression as well as dimensionality reduction, especially when the number of variables is large when compared to the observations. Although, PLS has superior predictability when compared to most linear regressors, recent research [39] has shown that the presence of a large number of variables introduces noise into the PLS latent scores. Hence, Sparse partial least squares (sPLS) regression has been proposed as an improvement to the conventional PLS. Motivated by its predictive power, dimensionality reduction and ability to filter irrelevant variables, we propose to use sPLS at the core of our invertible ageing framework. To the best of our knowledge, sPLS has not been explored before for age estimation.

Given $\boldsymbol{X}$ an $n \times p$ matrix of face features, and $\boldsymbol{Y}$ an $n \times q$ matrix of response variables, we can use PLS to express the predictor and response variables as linear equations given by,

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{C}^T + \boldsymbol{E}, \boldsymbol{Y} = \boldsymbol{T}\boldsymbol{D}^T + \boldsymbol{F}. \quad (1)$$

$\boldsymbol{T}$ an $n \times k$ matrix is called the score or latent matrix. Usually, $k \ll p$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are loadings, while $\boldsymbol{E}$ and $\boldsymbol{F}$ are matrices of residuals. In order to find $\boldsymbol{T}$, PLS produces a matrix of weights $\boldsymbol{W}$ such that,

$$\boldsymbol{T} = \boldsymbol{X}\boldsymbol{W}, \quad (2)$$

Using the SIMPLS algorithm [40], the kth weight is computed by,

$$\widehat{\boldsymbol{w}}_k = argmax_{\boldsymbol{w}}\, \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{w}, \quad (3)$$

such that $\boldsymbol{w}^T\boldsymbol{w} = 1$ and $\boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w}_i = 0$, for $i = 1 \ldots k-1$.

After estimating the scores $\boldsymbol{T}$, the PLS regression coefficient is computed by ordinary least squares minimization,

$$\beta^{PLS} = \boldsymbol{W}\boldsymbol{D}^T. \quad (4)$$

Hence, relationship between the predictor $\boldsymbol{X}$ and response variable $\boldsymbol{Y}$ is formulated as,

$$\boldsymbol{Y} = \boldsymbol{X}\beta^{PLS} + \boldsymbol{F}. \quad (5)$$

To ensure the selection of the most relevant variables, a form of sparsity can be integrated into the conventional PLS dimension reduction procedure. The technique which is termed sparse partial least squares regression (sPLS), has proven to be more efficient than PLS when the problem is ill posed. [39], i.e. for $n \ll p$.

As described in [39], we can derive the sPLS regression equation by reformulating the objective function in equation (3) by imposing LASSO L_1 regularisation on a surrogate direction vector $\boldsymbol{c}$ such that,

$$\min_{\boldsymbol{w},\boldsymbol{c}}\{-\kappa\boldsymbol{w}^T\boldsymbol{M}\boldsymbol{w} + (1-\kappa)(\boldsymbol{c}-\boldsymbol{w})^T\boldsymbol{M}(\boldsymbol{c}-\boldsymbol{w}) + \lambda_1|\boldsymbol{c}|_1 + \lambda_2|\boldsymbol{c}|_2\} \text{ such that } \boldsymbol{w}^T\boldsymbol{w} = 1, \quad (6)$$

where, $\boldsymbol{M} = \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}$. The L_1 penalty (λ_1) is used to ensure sparsity, while the L_2 tuning parameter (λ_2) is used to avoid singularity in $\boldsymbol{M}$. For a univariate predictor $\boldsymbol{Y}$, the solution above does not depend on κ and λ_2. Hence, only λ_1 needs to be tuned. Moreover, it does not have to be explicitly computed for each directional score. Instead, a soft threshold estimate has been formulated in the literature [39]. i.e.,

$$\widetilde{\boldsymbol{w}} = \left(|\widehat{\boldsymbol{w}}| - \eta \max_{1 \le i \le p} |\widehat{\boldsymbol{w}}_i|\right) \boldsymbol{I}_{(|\widehat{\boldsymbol{w}}| \ge \eta \max_{1 \le i \le p} |\widehat{r}_i|)} sign(\widehat{\boldsymbol{w}}). \quad (7)$$

A different η is required for each directional vector ($i = 1 \ldots k$); this is not computationally feasible. Hence, a single sparsity parameter is chosen within the range $0 \le \eta \le 1$. Eventually, we only have two parameters to tune the number of components k and the sparsity parameter η. It is worth noting that in this work, $\boldsymbol{Y}$ represents the ages of individuals, hence it is univariate ($q = 1$).

4. Experiments and Results

Here, we present the databases used for our experiments, data pre-processing procedure, analyses of the performance of the three (ResNet-50, ResNet-101, and ResNet-152) features, and finally, our results are compared to other state-of-the-art pre-trained CNN features. It is worth mentioning that just like the ResNet features we extracted, we used the output of last pooling layer for all the pretrained models to extract the image feature. Additional comparisons are made to the biologically inspired features (BIF) [41] which has been one of the most favoured algorithms in age estimation literature [42].

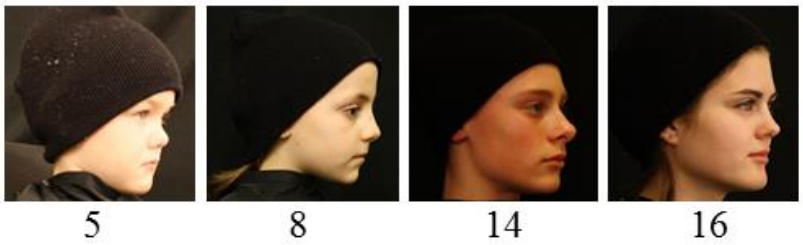
4.1. Data

One major challenge of using profile face images for age estimation is the lack of sufficient labelled data that can be used for both training and testing. Apparently, existing benchmark ageing datasets such as FGNET and the Morph II do not have extreme-view (profile) face images. Fortunately, we were able to conduct our experiments on the under listed controlled, semi-controlled and unconstrained datasets.

4.1.1 HQfaces: This is part of the Siblings database [43] collected by Politecnico di Torino's Computer Graphics and Vision group. The database has both frontal (HQf) and profile (HQfp) images, all having resolution of 4256×2832. The subjects are Caucasian with ages between 3 and 50 years. In this work, all 158 profile images with neutral expressions, provided in HQfp subset of the HQfaces database were used. Fig. 2(a) shows samples of the HQfp faces.

4.1.2 Dartmouth database of children's faces (DDCF): DDCF [44] contains front and profile images of 80 Caucasian children, with a gender ratio of 1:1 and ages between 5 and 16 years (See Fig. 2(b) for example of these images). This database was formed at the University of Dartmouth, USA. The images used for our experiment display neutral facial expression.

a



b

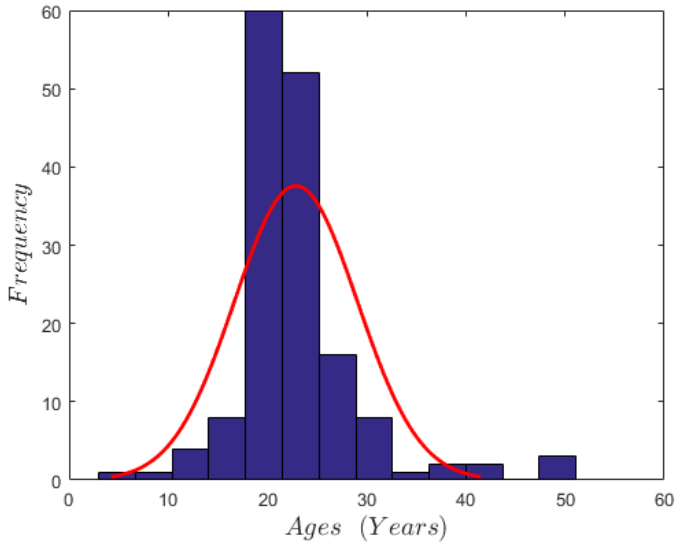
c

d

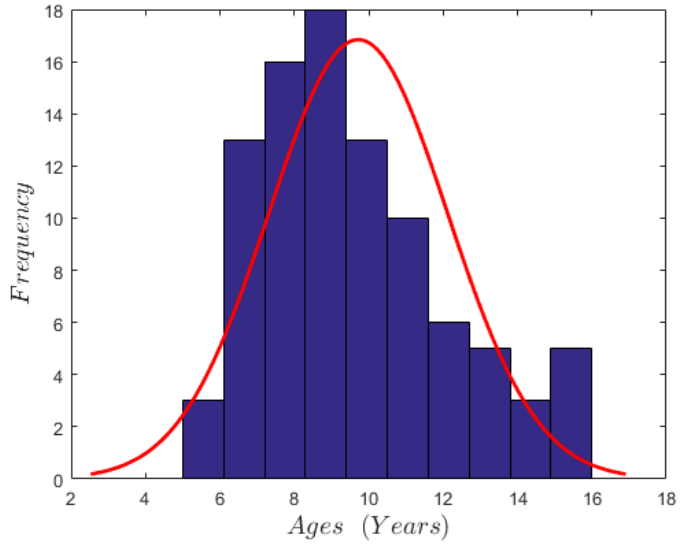
Fig. 2. *Sample of face images used for our experiments (a) HQfaces profile images (HQfp) (b) DDCF images (c)Color FERET Database (d) Unconstrained Images of Celebrities.*

4.1.3 Color FERET Face Database: The FERET database [45] is a publicly available dataset that is commonly used for evaluating face recognition, and other face related algorithms. FERET face images were collected under semi-controlled environment between 1993 and 1996. Thus, images in the dataset display variations in scale, expression, pose and illumination, additionally, some subjects were photographed having their pairs of glasses. In this work, left and right profile images retrieved from *pr* and *pl* subsets of the album were used. These jointly form a total of 2674 images of 994 subjects, with ages ranging between 10 and 70. A sample of the FERET face images used is shown in Fig. 2(c).

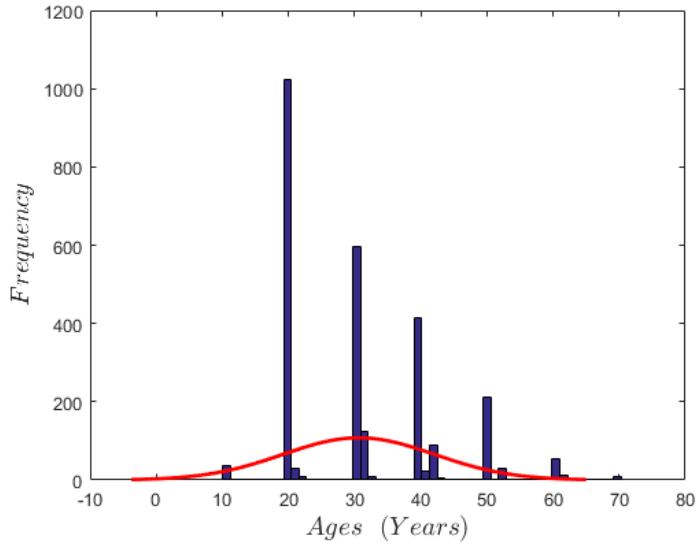
4.1.4 Unconstrained Images Database: To further evaluate the algorithm proposed in this work, a dataset of 700 uncontrolled images of 200 unique individuals was formed from the Internet. These subjects are mainly well known people, with ages ranging from 2 to 75 years; their real ages were deduced from online image captions, and dates of important events on which the pictures were taken. This dataset has varying picture qualities, from grayscale to colour images, with diverse illumination, resolution, head pose, facial expression and background clutter. In total, the database has a male to female ratio of 4:3. Sample images from this dataset are shown in Fig. 2(d). Age distributions of the four datasets mentioned above are presented in Fig. 3.



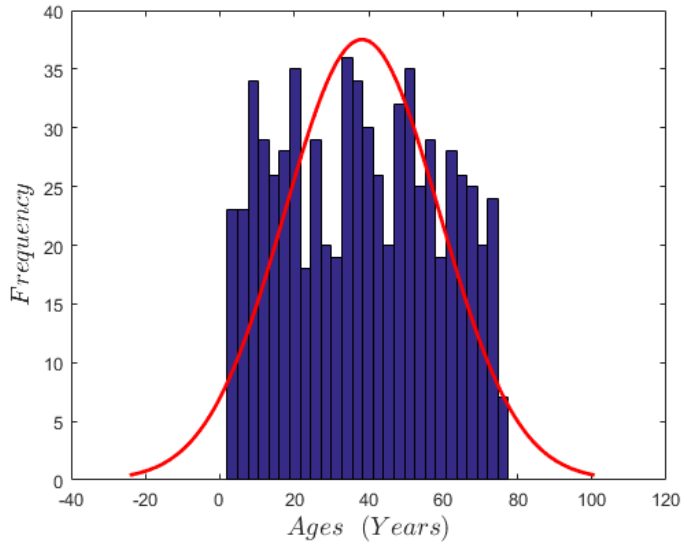
a



b



c



d

Fig. 3. *Age distribution (a) HQfaces profile images (HQfp) (b) DDCF images (c)Color FERET Database (d) Unconstrained Images of Celebrities.*

4.2. Data Pre-processing

All images used in our experiments were first resized to 224 x 224 pixels; this was done to get them to conform to ResNet's input size. Furthermore, we augmented the training data using random crops, zoom, and rotations. This eventually led to a six fold increase in data size.

4.3. Estimation and Evaluation

Utilising features from ResNet-50, ResNet-101, and ResNet-152 models, three sets of estimation were conducted on each dataset; this entails feeding facial features into the sPLS regressor. For regression, we chose the optimum number of sPLS latent scores (k), and the sparsity parameter η, using cross validation.

In line with earlier works on age estimation [42], we employed the leave-one-person-out (LOPO) cross validation protocol. Hence, for each database, the sPLS regressor is iteratively trained with the images of $n - 1$ subjects, i.e. all but 1 person. The estimator is then tested on images of the subject that wre left out during training. Finally, the results are averaged over all n folds.

For performance evaluation, we initially used Mean Absolute Error (MAE) and Cumulative Score (CS), both expressed as,

$$MAE = \sum_{i=1}^{N} | y' - y | / N, \tag{8}$$

$$CS(m) = N_{abs\_error \leq m} / N \times 100\%. \tag{9}$$

where y' is the estimated age, and the ground truth age is y, N the number of test images, and $N_{abs\_error \leq m}$ is the number of images on which the system makes absolute error less than or equal to m years.

Table 1 Comparison of ResNet features on HQfaces [43]

Features	MAE	CS < 10yrs
ResNet-50	4.09	93.07%
ResNet-101	4.00	93.97%
ResNet-152	3.74	95.93%

Table 2 Comparison of ResNet features on DDCF [44]

Features	MAE	CS < 10yrs
ResNet-50	1.28	100%
ResNet-101	1.25	100%
ResNet-152	1.20	100%

Table 3 Comparison of ResNet features on FERET [45]

Features	MAE	CS < 10yrs
ResNet-50	5.67	84.35%
ResNet-101	5.61	84.66%
ResNet-152	5.50	84.92%

Table 4 Comparison of ResNet features on Unconstrained Images

Features	MAE	CS < 10yrs
ResNet-50	6.58	79.90%
ResNet-101	6.51	80.11%
ResNet-152	6.37	80.42%

Tables 1, 2, 3 and 4 above, show that ResNet-152 features give the best result on all four datasets, by achieving MAEs of 3.74, 1.20, 5.50 and 6.37 on HQfaces, DDCF, FERET, and the unconstrained datasets respectively. The cumulative score (CS) also reveal performance improvement as the depth increased. We observed a staggering performance in the two controlled datasets; all the data in DDCF, and up to 95% of HQfaces images had estimation errors of less than 10 years. Additionally, it is also obvious that ResNet-101 consistently outperforms ResNet-50, this clearly corroborates the findings of [27] i.e. as ResNet layers get deeper, the performance gets better.

Next, we compared the performance of our best result with state-of-the-art algorithms. To be precise we compared to features extracted using the BIF [41], and three other Pretrained ConvNets; AlexNet [46], VGGNet-16 [47], and GoogleNet [48]. To ensure fairness, comparison experiments were conducted using the same variable selection procedure that we employed for the ResNet features, hence, the LOPO approach was utilised for all the test cases. As can be seen in Tables 5, 6, 7, 8 as well as Fig. 4, the ResNet features consistently outperform the other algorithms.

Table 5 Comparison to state-of-the-art algorithms on HQfaces [43]

Features	MAE	CS < 10yrs
BIF	5.30	87.23%
AlexNet	5.07	89.34%
VGGNet-16	4.49	90.50%
GoogleNet	4.28	91.14%
ResNet-152	**3.74**	**95.93%**

Table 6 Comparison to state-of-the-art algorithms on DDCF [44]

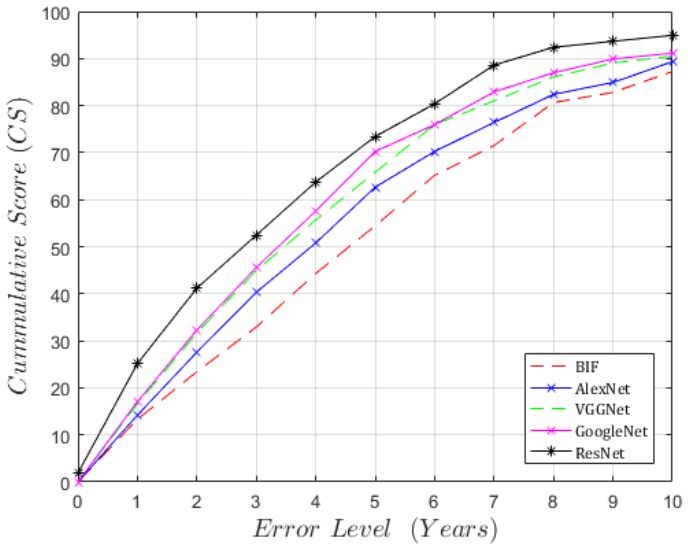
Features	MAE	CS < 10yrs
BIF	1.61	97.91%
AlexNet	1.57	98.21%
VGGNet-16	1.48	98.90%
GoogleNet	1.35	100%
ResNet-152	**1.20**	**100%**

Table 7 Comparison to state-of-the-art algorithms on FERET [45]

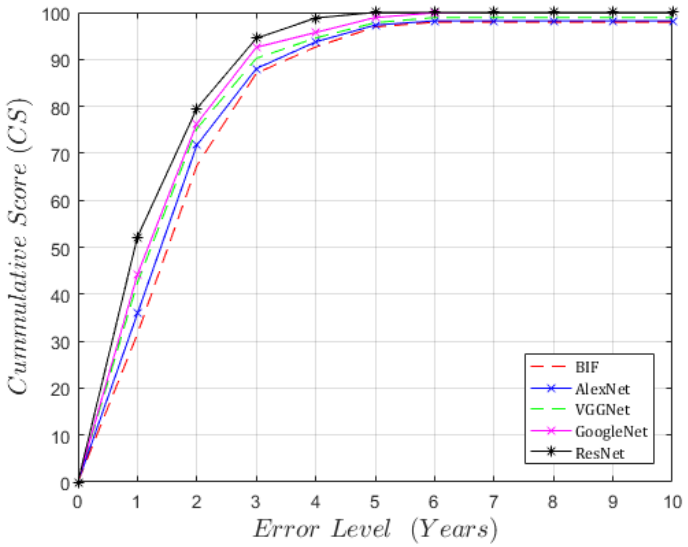
Features	MAE	CS < 10yrs
BIF	6.26	79.94%
AlexNet	6.00	81.13%
VGGNet-16	5.78	83.42%
GoogleNet	5.70	84.23%
ResNet-152	**5.50**	**84.92%**

Table 8 Comparison to state-of-the-art algorithms on Unconstrained Images

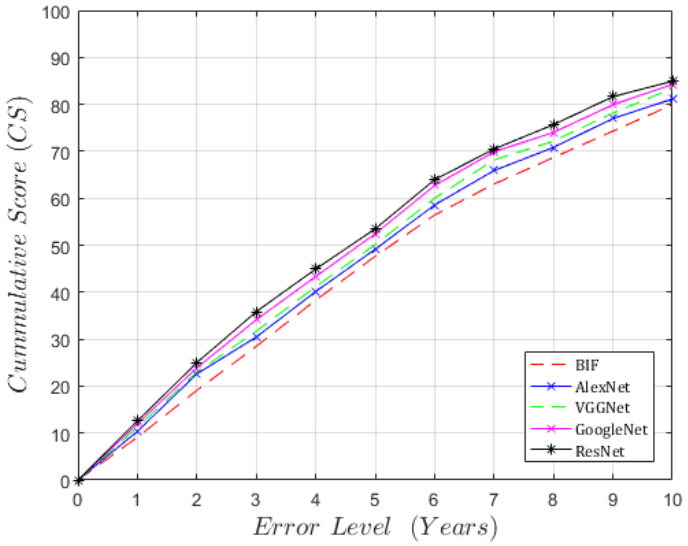
Features	MAE	CS < 10yrs
BIF	7.47	75.44%
AlexNet	7.15	76.63%
VGGNet-16	7.01	78.92%
GoogleNet	6.59	79.74%
ResNet-152	**6.37**	**80.92%**



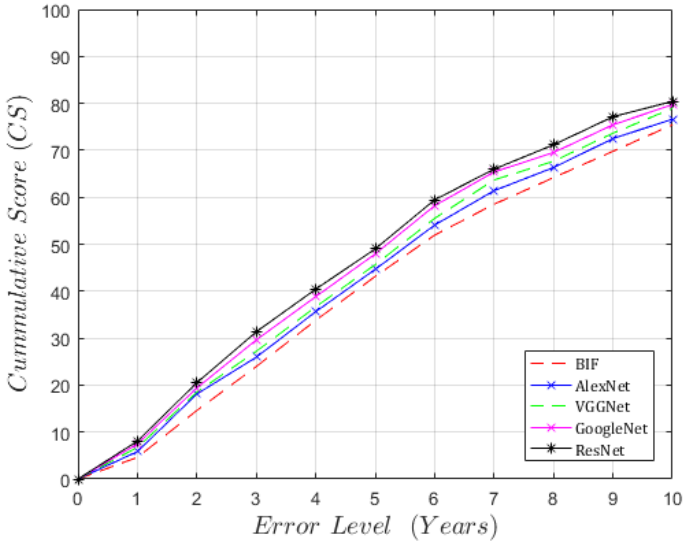
a



b



c



d

Fig. 4. *Comparison of CS (a) HQfaces profile images (HQfp) (b) DDCF images (c)Color FERET Database (d) Unconstrained Images of Celebrities.*

5. Conclusion

In this paper, we presented a simple yet powerful method for automatic age estimation. To the best of our knowledge, this is the first attempt to estimating ages from face profile images. While most researchers focussed on frontal and/or semi frontal images, real world applications present us with diverse head poses and face-views. Deep learned, residual neural network (ResNet) features were used for face representation, the extracted features were then fed into a sparse partial least squares regressor to achieve age estimation. Results show that ResNet features encode much age related facial features and a comparison of the proposed method to state-of-the-art algorithms reveals the superiority, generality, and effectiveness of deeply learned ResNet representations. Our results further confirm the findings of He *et al.* [27]; the performance of ResNets get better as the network depth increases. Thus, we conclude that using deep learning, automatic age estimation from profile-view face images is feasible. We hope that researchers will direct their attention towards this new problem.

However, our work does leave room for future improvements. For example, the performance degrades when evaluated on semi-controlled and unconstrained images. Additionally, the size of the unconstrained images' dataset is also relatively small. In the future, age estimation results could therefore be improved by implementing more sophisticated feature extraction techniques, or by implementing an end-to-end deep learning architecture. The unconstrained images dataset we used for our experiments will made publicly available to research community, we also intend to collect more unconstrained images to make the dataset bigger.

6. References

[1] M. Kaur, R. K. Garg, and S. Singla, "Analysis of facial soft tissue changes with aging and their effects on facial morphology: A forensic perspective," *Egypt. J. Forensic Sci.*, vol. 5, no. 2, pp. 46–56, 2015.

[2] Y. Fu, G. Guo, and T. Huang, "Age Synthesis and Estimation via Faces : A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, 2010.

[3] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output cnn for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4920–4928.

[4] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognit.*, vol. 44, no. 6, pp. 1262–1281, Jun. 2011.

[5] C. Fernandez, I. Huerta, and A. Prati, "A Comparative Evaluation of Regression Learning Algorithms for Facial Age Estimation," *FFER conjunction with ICPR, Press. IEEE*, 2014.

[6] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *Inf. Forensics Secur. IEEE Trans.*, vol. 9, no. 12, pp. 2170–2179, 2014.

[7] J. Kim, D. Han, S. Sohn, and J. Kim, "Facial age estimation via extended curvature Gabor filter," in *Image Processing (ICIP), 2015 IEEE International Conference on*, 2015, pp. 1165–1169.

[8] R. Weng, J. Lu, G. Yang, and Y.-P. Tan, "Multi-feature ordinal ranking for facial age estimation," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, 2013, pp. 1–6.

[9] H. Han, C. Otto, A. K. Jain, and E. Lansing, "Age Estimation from Face Images : Human vs . Machine Performance," 2013.

[10] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 534–541.

[11] T. Liu, J. Wan, T. Yu, Z. Lei, and S. Z. Li, "Age Estimation Based on Multi-Region Convolutional Neural Network," in *Chinese Conference on Biometric Recognition*, 2016, pp. 186–194.

[12] Y. H. Kwon and V. Lobo, "Age classification from facial images," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR-94*, pp. 762–767, 1994.

[13] N. Ramanathan and R. Chellappa, "Modeling Age Progression in Young Faces," *2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 1*, vol. 1, pp. 387–394, 2006.

[14] I. Engineering, "3D Age Progression Prediction in Children ' s Faces with a Small Exemplar-Image Set *," vol. 1148, no. 100, pp. 1131–1148, 2014.

[15] A. Gunay and V. V Nabiyev, "Automatic detection of anthropometric features from facial images," in *2007 IEEE 15th Signal Processing and Communications Applications*, 2007, pp. 1–4.

[16] A. Lanitis, C. Taylor, and T. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, 2002.

[17] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," *Proc. 14th Annu. ACM Int. Conf. Multimed. - Multimed. '06*, p. 307, 2006.

[18] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning Auto-Structured Regressor from Uncertain Nonnegative Labels," *2007 IEEE 11th Int. Conf. Comput. Vis.*, pp. 1–8, 2007.

[19] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 1383–1386.

[20] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 112–119.

[21] M. Y. El Dib and M. El-Saban, "Human age estimation using enhanced bio-inspired features (EBIF)," *2010 IEEE Int. Conf. Image Process.*, pp. 1589–1592, Sep. 2010.

[22] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 10–15.

[23] X. Liu, S. Li, M. Kan, J. Zhang, S. Wu, W. Liu, H. Han, S. Shan, and X. Chen, "Agenet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 16–24.
[24] Y. Zhu, Y. Li, G. Mu, and G. Guo, "A study on apparent age estimation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 25–31.
[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv Prepr. arXiv1512.03385*, 2015.
[28] H. Azizpour, A. Sharif Razavian, J. Sullivan, A. Maki, and S. Carlsson, "From generic to specific deep representations for visual recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 36–45.
[29] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717–1724.
[30] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
[31] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806–813.
[32] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv Prepr. arXiv1312.6229*, 2013.
[33] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
[34] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European Conference on Computer Vision*, 2014, pp. 584–599.
[35] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
[36] B. Athiwaratkun and K. Kang, "Feature Representation in Convolutional Neural Networks," *arXiv Prepr. arXiv1507.02313*, 2015.
[37] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv Prepr. arXiv1505.00387*, 2015.
[38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[39] H. Chun and S. Keleş, "Sparse partial least squares regression for simultaneous dimension reduction and variable selection," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 72, no. 1, pp. 3–25, 2010.
[40] S. De Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemom. Intell. Lab. Syst.*, vol. 18, no. 3, pp. 251–263, 1993.
[41] T. S. Huang, "Human age estimation using bio-inspired features," *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 112–119, Jun. 2009.
[42] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the FG-NET ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, 2016.
[43] T. F. Vieira, A. Bottino, A. Laurentini, and M. De Simone, "Detecting siblings in image pairs," *Vis. Comput.*, vol. 30, no. 12, pp. 1333–1345, 2014.
[44] K. A. Dalrymple, J. Gomez, and B. Duchaine, "The dartmouth database of children's faces: Acquisition and validation of a new face stimulus set," *PLoS One*, vol. 8, no. 11, pp. 1–7, 2013.
[45] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.
[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.