Deeply-Learned Feature for Age Estimation

Xiaolong Wang[1], Rui Guo[2], Chandra Kambhamettu[1]
[1]Department of Computer and Information Sciences, University of Delaware, Newark, DE, USA.
[2]University of Tennessee, Knoxville, USA.
[1]{xiaolong, chandrak}@udel.edu, [2]rguo1@utk.edu

Abstract

Human age provides key demographic information. It is also considered an important soft biometric trait for human identification or search. Compared to other pattern recognition problems (e.g., object classification, scene categorization), age estimation is much more challenging since the difference between facial images with age variations can be more subtle and the process of aging varies greatly among different individuals. In this work, we investigate deep learning techniques for age estimation based on the convolutional neural network (CNN). A new framework for age feature extraction based on the deep learning model is built. Compared to previous models based on CNN, we use feature maps obtained in different layers for our estimation work instead of using the feature obtained at the top layer. Additionally, a manifold learning algorithm is incorporated in the proposed scheme and this improves the performance significantly. Furthermore, we also evaluate different classification and regression schemes in estimating age using the deep learned aging pattern (DLA). To the best of our knowledge, this is the first time that deep learning technique is introduced and applied to solve the age estimation problem. Experimental results on two datasets show that the proposed approach is significantly better than the state-of-the-art.

1. Introduction

A recent study in the field of Psychology [36] has shown that the proportions and expressions of human faces are fundamental to human social interaction. For example, the face is essential in identifying origins, emotional tendencies, health qualities, and social information. Lots of visually nonverbal information is conveyed by the human face, including identity, age, gender, ethnicity, emotion, etc. Faces are essential for daily communication.

Facial attributes play significant roles in the application of image processing technology, such as human-computer interaction (HCI). It is known that one important criterion

Figure 1. An illustration of the aging process for two different individuals. Each row shows images of the same person at different ages.

to evaluate the performance of these systems is their capabilities to correctly interpret facial images in real time. Acquiring these facial attributes accurately from facial images should be a great help for improving these systems' reacting and performing functions. Automatic age estimation systems have a wide range of applications. In the security area, a well designed Age Specific Human-Computer Interaction system (ASHCI) can help prevent minors browsing adult web pages or purchasing age restricted material from vending machine. ASHCI can also facilitate our daily life, such as an ASHCI installed vehicle can prevent children from starting the engine without the guidance of adults. An ASHCI system can also provide a warning when the driver leaves the kids alone in the vehicle without taking any protective measures. Age information can also be used in law enforcement. It can be used to quickly locate the suspects in a specific age group by limiting the searching range in the gallery set. This can improve the efficiency of matching. Age estimation systems are also useful in the commercial domain. For instance, store owners can adjust decoration styles and advertisements according to their customers' demographics.

Although age estimation based on images is an important technique in real-world applications, it is still a challenging

978-1-4799-6683-7/15 $31.00 © 2015 IEEE
DOI 10.1109/WACV.2015.77

problem. As illustrated in Fig. 1, we can see that the aging process of different individuals varies greatly. As discussed in [16, 18], the process of aging is very complicated. Many factors affect the aging process. In general, these factors can be divided into two different categories, which are internal and external factors. The internal factors are mainly determined by physiological elements, such as genes. External factors include living environment, health condition, lifestyle, etc. Extracting robust aging features invariant to the influence brought by individual differences is still an open problem.

In this paper, we explore a new framework to learn aging features directly from the data. The learned aging features are expected to be more general and discriminative than previous pre-defined feature extraction algorithms [15, 18, 13]. This work has several novelties. First, we propose a novel method for automatic age estimation. The proposed scheme is based on a deep learning model (CNN), as far as we know, this is the initial work that CNN is applied to the age estimation problem. Second, to evaluate the performance of the proposed scheme, we compare with the most recent approaches on two benchmark datasets. Experimental results demonstrate a significant improvement compared with the state-of-the-art. Third, we adopt manifold learning approaches with this scheme, and an evaluation for different manifold learning approaches is given. Furthermore, different regression and classification approaches are evaluated based on the aging features obtained from the deep learned neural network.

2. Related Work

In the past years, many efforts have been devoted to the human image based age estimation study. These works can be divided into two subproblems: how to extract aging feature, and how to predict the age based on the extracted feature. For aging feature representation, one representative work used anthropometric shape, such as the work conducted by Kwon and Lobo [29]. They used anthropometric information to predict age. Their method is based on craniofacial development theory and analysis of skin wrinkles. A similar idea is also adopted in [37], where a craniofacial growth model was proposed to characterize the growth shape variations. The problem of these works is that these methods can only be used for coarse age estimation, such as to estimate a general age group for a probe image, whether it belongs to young adult, senior adult or babies [29]. This kind of work were not designed for continuous age classification, which limits the application range in the problem of age estimation. Guo et al. adopted Bio-inspired feature [40] for age estimation work in [23, 24], where several Gabor filters with different orientations and scales are predefined to extract age feature from the facial image.

Another representative work for age image representation uses aging pattern subspace [15]. The basic idea is to make use of sorted individual's facial images to construct the aging subspace for age estimation. The age label of the test face is predicted based on the projection in the subspace which is used to reconstruct the facial image. One advantage of this kind of work is that the constructed learning subspace can handle missing ages in the sequence. Manifold learning is also used in extracting aging features [14, 35]. Motivated by the ordinal characteristic of aging process, such as an 8 year old's face is much more closely related to the face of 11 year old's, the ranking scheme was proposed to solve age estimation [9, 35].

The age estimation problem can be regarded as a classification [15] or regression problem [30, 46, 14]. Recently, several new regression approaches have been introduced for age estimation. Chen et al. adopted the idea of cumulative attributes into the regression problem used for age estimation [10]. Guo and Mu used partial least squares (PLS) [44] regression and canonical correlation analysis (CCA) [25] to learn aging pattern combined with gender and ethnicity in [22].

The structure of our proposed work is organized as follows. The proposed approach is illustrated in Section 3, where we give a detail description about the proposed scheme. In Section 4, we talk about the dataset and experimental setting. Experimental evaluations of the proposed approach and the comparison with previous methods are discussed in Section 4. The final conclusion is given in Section 5.

3. Proposed Approach

We first introduce the fundamentals of CNN, then talk about our feature extraction scheme. A brief introduction of Convolutional Neural Network is given, mainly focusing on why the learned "deep" feature are effective in characterizing age patterns and how to make use of the learned feature in age estimation problem. Then we investigate different manifold learning approaches in age estimation for its good performance to capture the underlying aging structure and represent age information in a low dimensionality. We also investigate several popular classification and regression methods including Support Vector Regression (SVR) [43], Support Vector Machines (SVMs) [5], PLS and CCA for age estimation.

3.1. Convolutional Neural Network (CNN)

To learn aging patterns from various facial images, we need a model with a powerful learning capacity. Recently, deep learning models have been used to extract multiple layers of features in a hierarchical structure. The constructed structure can capture hierarchical, abstract, and translation invariant features. Deep learning models have demonstrated very promising and plausible results in many applications

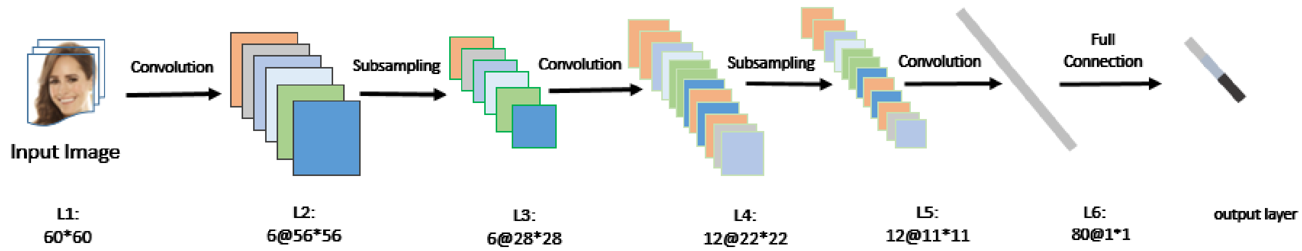

Figure 2. Illustration of CNN structure for extracting deep learned age features.

[26, 2, 33, 27, 28, 8, 51]. From these works, we find that deep learning methods are good at characterizing the high-level abstractions of visual data by using a deep architecture, which is composed of multiple non-linear transformations. Among them, Convolutional Neural Network (CNN) [32] has demonstrated an excellent ability in capturing image characteristics [33, 27, 28]. CNN can be classified as a class of biologically-inspired, multi-layered neural network. It uses a feed-forward neural network. The construction of CNN was inspired by the biological processes designed to use multi-layer transformations to simulate human visual perception. The learned individual neurons in the feature map respond to overlapping regions in the visual field.

CNN has been widely used in many computer vision topics and obtains very promising results, such as in visual object recognition, detection and image retrieval [28, 12, 11, 49]. Especially, seen from the visualization of learned features in [49], CNN has shown to be very effective in learning interpretable features at different layers. Benefiting from its good ability in extracting features from images, in this work, we apply CNN to extract aging feature from static facial images. We reason that the learned weights from the deeper layers are more specific to the images of the training dataset for age estimation problem. The typical CNN structure is composed of convolution layer and sub-sampling (pooling) layer, which are stacked alternatively. At the top of the network, a fully connected MLP (multiple logistic perception) is applied.

3.1.1 Convolution Layers

Each neuron in the convolution layer is formed using the input from a local receptive field in the preceding layer and the learned kernels (weights). Neurons within the same feature map share the same kernels but are obtained using different input receptive fields. The kernels used for different feature maps in the same layer are different. Subsequently, an activation function is used. This can be represented as:

$$u_j^l = f(\sum_{i=1}^{N} u_i^{l-1} * w_{ji} + b_j^l), \tag{1}$$

where u_i^{l-1} is the input neuron from $l-1$ layer, N is the total number of input neurons, b_j^l denotes the bias, f is the activation function. In this work, logistic (sigmoid) function is used as the activation function.

3.1.2 Sub-sampling Layers

The operation in the sub-sampling layer is conducted by the downsampling of given input feature maps. It only changes the size of the input maps while not altering the number of input maps. There are several ways to implement the sub-sampling operations, such as averaging or taking the maximum, or using learned combinations of the neurons in the block. In this work, average pooling is applied.

The sub-sampling operation is used to maintain specificity. Although the concept seems simple, it is efficient in characterizing the feature for specific objects. This concept has adopted schemes in the mammalian visual cortex [41]. Its learned architecture has been proven to be very efficient in learning hierarchical features for the specific problem.

3.1.3 Training Process

In general, the goal of training CNN is to minimize the error function. For a classification problem with N training samples, the reconstruction error (sensitivity) is calculated by

$$E = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}(s_k^n - y_k^n)^2, \tag{2}$$

where s_k^n represents the k-th dimension of the n-th sample's label, y_k^n is the output value of the k-th output layer unit corresponding to the n-th input pattern. From Eqn. 2, we can see the error of the whole dataset as the sum over the individual errors on each neuron pattern. Thus the error for a single pattern can be formatted as $E^n = \frac{1}{2}\sum_{k=1}^{K}(s_k^n - y_k^n)^2$. During the training stage of CNN, there are two schemes applied, which are feedforward pass and backprogration pass. In feedforward pass, feature maps are transformed from one layer to the successive layer via pre-defined activation functions with learned parameters (weights and basis). For example, the output of layer l is defined as $v^l = f(u^l)$, where $u^l = (W^l v^{l-1} + b^l)$. During the backprogration pass, weight W^l and b^l are updated via stochastic gradient descent, which is trained on the backpropagating the derivative of the loss (errors) obtained in the feedforward pass with respect to the parameters throughout the network. Normalizing the data to have normal distribution in the feature space can accelerate the convergence [34]. For more mathematical theory, we refer to [3, 4] for specific details.

3.2. Deep Learned Aging Pattern (DLA)

Our framework for age estimation contains several steps, which are illustrated in Fig. 2. We train the deep model of CNN for a multi-classes age estimation task. The leaned structure is used to predict the age given a testing facial image. We call the feature based on the proposed scheme deep learned aging pattern (DLA).

Our whole network is comprised of 6 layers. At the input level, $2D$ grayscale face images of size 60 by 60 pixels are used as input to the first convolutional layer with the kernel size of 5×5. After applying the filter with size $a \times b$ to a feature map of $h \times w$, the size of output is $(h - a + 1) \times (h - b + 1)$. So the size of feature maps in layer L_2 after convolution is 56×56. At the end of the layer, an activation function (logistic function) is applied to obtain feature maps. Within a feature map, all the units share the same set of weights for the filter. The obtained feature maps in the convolution layer go through a pooling layer.

The L_3 layer is obtained by pooling to the feature map obtained in layer L_2. We apply 2×2 subsampling on each of feature maps in the L_2 layer. This leads to the reduction of the spatial resolution while keeping the same number of feature maps. As analyzed in [28], pooling layers are used to "filter" the convolution networks' output. This operation is very robust to the local translation and rotations. The next convolution layer $L4$ is obtained by the convolution with a kernel size of 7×7 on the feature maps from layer L_3. The next layer L_5 is also obtained by the pooling operation of each feature maps from layer L_4. The L_6 layer consists of 80 feature maps of size 1×1. Each unit is connected to all feature maps in layer L_5. The output layer (softmax classifier) is fully connected to L_6 layer. Feature maps obtained in these layers are constructed to extract low-level features, including edges and texture. During the training stage, face images with age labels are used for the supervised learning. These filters are used to extract those features which can characterize the age information. A small neighborhood in one layer is connected to the units in the successive layer. This adopts the idea of extracting feature from local receptive fields. Another advantage of applying CNN in this work is filter weights are shared by the units in the same feature map. This can greatly reduce the number of parameters for training.

Most works usually use the output of the top layer in a deep learning model as the feature representation in their problem [38, 17]. However, unlike these works, in this paper, motivated by the idea in [49], where representations in different layers respond to particular activations, we investigate the use of these extracted features in the problem of age estimation and whether combing these features is an effective way for predicting age. As far as we know, this is the first time that deep learned features are utilized in such a way for age estimation.

After training the CNN, we can extract the feature from different layers. We find that the dimensionality of features in CNN layers is high, such as layer L_4 has a dimension as high as 5808 by itself. With the consideration of the simplicity and efficiency, principal component analysis (PCA) [42] is applied for feature dimension reduction. Assuming $X_l = [x_1, \ldots, x_N]$ to be the feature vectors of N samples with the dimension T extracted from the $l\text{-}th$ layer. PCA aims to find the projection matrix P, which projects X_l into the new subspace Y_l based on $Y_l = P^T X_l$, where $Y_l = [y_1, \ldots, y_N]$ with dimension t, and $t < T$. P satisfies $P = argmin\, P^T CP$, where $||P|| = 1$. C is the covariance matrix, $C = \sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T$, $\bar{x}$ is the mean vector obtained from training samples. Based on the experimental results, we found that the age estimation error nearly does not change when PCA is applied. This step is very efficient in speeding up training of the age estimation model.

Afterwards, features extracted from different layers are concatenated together to obtain our aging pattern. This can be represented as $F = [Y_{L_i}, \ldots, Y_{L_J}]$, where Y_{L_J} indicates the feature extracted from layer L_J. In this work, the feature maps extracted from layer 2 to layer 5 are concatenated together as the aging feature (DLA).

Manifold learning has been applied in this work for its good performance in learning aging patterns and capturing the underlying face aging structure [14]. . Given the feature vector F with the size of $d \times N$, where N is the number of training samples and d is the number of feature dimensions. The goal of manifold learning is to find the projection matrix M with dimension $d \times k$ which satisfiies $Y = M^T F$, where $k < d$. In this work, three different manifold learning schemes are analyzed and compared. They are Marginal Fisher Analysis (MFA) [48], Orthogonal Locality Preserving Projections (OLPP) [6] and Locality Sensitive Discriminant Analysis (LSDA) [7].

Table 1. Illustration of the used datasets. N_{in} indicates the number of individuals. Age_r represents the age range of the corresponding dataset.

Dataset	N_{in}	Age_r
MORPH [39]	5475	16-77
FG-NET[1]	1002	0-69

3.3. Regression or Classification

As discussed in [19, 31], age estimation can be classified as a regression problem, since each age can be considered as a regression value. Meanwhile, the age estimation problem can be categorized as a classification problem, where we can consider each age as a class label. Since it is the first time using a deep learned feature for age estimation, it remains unknown which scheme performs better. In this work, both SVMs for age classification and SVR for age regression are used and evaluated. PLS and CCA are also used as regression algorithms in this paper for their good performance in recent work [22].

4. Experiments

To evaluate the performance of the proposed scheme in age estimation, two different measurements are used in this work, which are Mean Absolute Error (MAE) and the Cumulative Score (CS). The MAE is calculated using the average of the absolute errors between the estimated result and the ground truth. MAE is calculated as

$$MAE = \frac{1}{K}\sum_{k=1}^{K}|g_k - g'_k|, \tag{3}$$

where g_k and g'_k represent the ground truth age and estimated age respectively. K is the total number of test images. The cumulative score (CS) is represented as follows:

$$CS_a = \frac{K'_a}{K} \times 100\%, \tag{4}$$

where K'_a represents the number of probe facial images whose absolute error between the estimated age and the ground truth age is not greater than a years.

4.1. Datasets and Experimental Settings

In this work, to compare with state-of-the-art approaches and evaluate the proposed method, our experiments are conducted on two available public datasets, MORPH database [39] and FG-NET [1]. To make a fair comparison with the state-of-the-art methods, we adopt the same experimental setting as the works in [18, 10, 35]. The illustration of used datasets is given in Table 1.

Table 2. Comparison of different methods for age estimation on two benchmarks (measured in MAE).

Method	FG-NET [1]	MORPH [39]
AGES. [15]	6.77	8.83
RUN [47]	5.78	-
Ranking [46]	5.33	-
LARR [18]	5.07	-
SVR [18]	5.66	5.77
MTWGP [50]	4.83	6.28
OHRank [9]	4.85	5.69
PLO [35]	4.82	-
CA-SVR [10]	4.67	5.88
Proposed Scheme	4.26	4.77

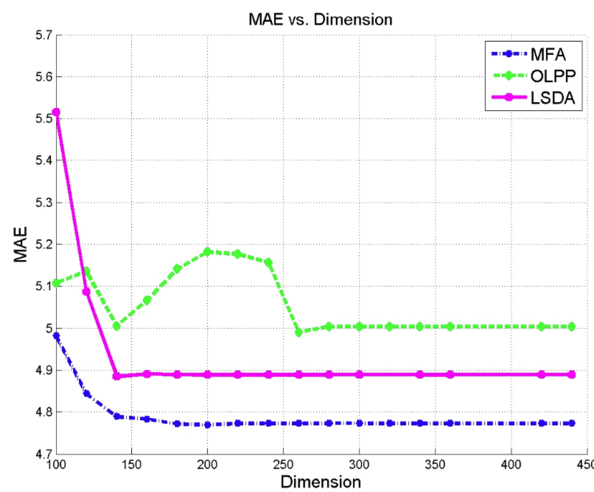

Figure 3. Comparisons of different manifold learning algorithms in terms of MAE for age estimation with different dimensions in MORPH dataset.

Following the experimental setting in [9, 10], for MORPH database, we randomly divide the whole dataset into two parts, one part is used for training, and the other one is used for testing. 80% data is used for training and 20% is used for testing. There is no overlap between the training and testing sets.

For FG-NET, leave-one person-out (LOPO) setting is used [47, 18, 9, 10]. This is determined by the limited size of FG-NET. FG-NET only includes 1002 facial images belonging to 82 subjects. However, the age range in this dataset is wide, ranging from 0 to 69 years old. This is the major reason for its popularity in age estimation work.

4.2. Experimental Results and Analysis

The experimental results are illustrated in Table 2 and Fig. 3. From the results, the proposed scheme demonstrates the effectiveness and robustness in age estimation. In con-

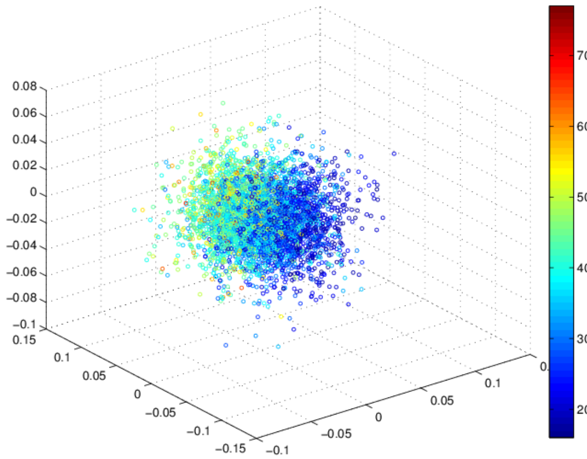

Figure 4. Visualization of the first three components of the extracted deep learned feature after manifold learning (MFA) on the MORPH dataset. Different colors represent different age categories.

trast, the new representation of aging feature outperforms other state-of-the-art approaches [15, 47, 46, 18, 50, 35, 10] by a large margin.

We can see that MAE is reduced by a large degree. The MAE obtained on MORPH is 4.77 compared to the current best result 5.69 achieved in [9]. The MAE from FG-NET is 4.26. This is less than 4.67 obtained in [10], which is the state-of-the-art approach. For the MAE reduction, the improvement is very evident. Assuming that the result obtained by [18] as the baseline, our approach outperforms the baseline by a full year ($|5.77 - 4.77| = 1.00$). This also indicates the deep learned aging feature is an effective representation for characterizing facial aging.

Based on the applied manifold learning algorithms, we can categorize the proposed schemes into three different methods, which are DLA+MFA, DLA+OLPP, DLA+LSDA. In general, discriminative mapping of the deep learned feature proved to be effective in learning the age function, especially using MFA as indicated in Fig. 3. It performs better than OLPP and LSDA. To show the distribution of the manifold learning results and give a better understanding of the learned feature, the first three feature vectors after projection are used and the samples are displayed in Fig. 4, which indicates that the extracted features are discriminative. In this work, three different manifold learning algorithms are used to provide a general framework for learning discriminative aging patterns. All of them have proved powerful in mapping deep learned aging pattern (DLA) into a new subspace. The regression scheme applied in this experiment is SVR. Another experiment is conducted to validate the effectiveness of the manifold learning.

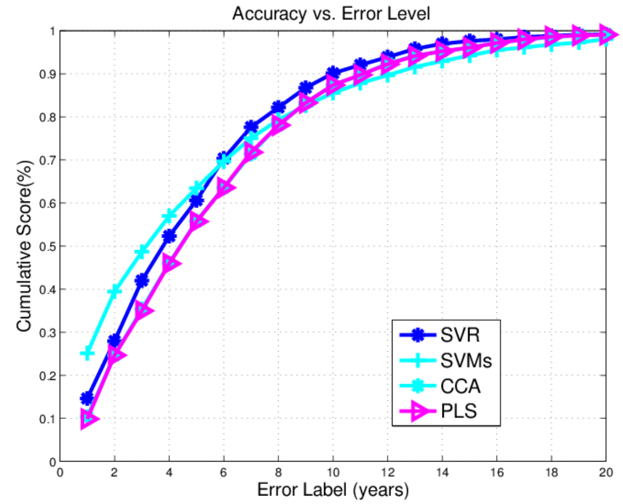

Figure 5. Cumulative scores of the algorithms with different settings on MORPH dataset.

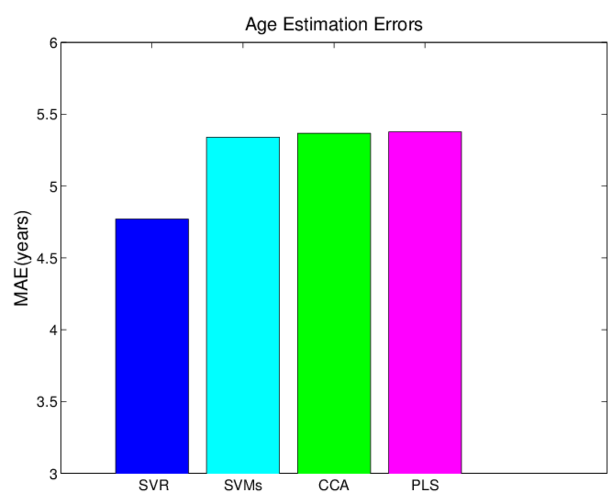

Figure 6. The age estimation errors using different methods on MORPH dataset.

ing. Without using the manifold learning, the MAE obtained on MORPH is 5.02, which is higher than 4.77. This shows the effectiveness of the manifold learning. Instead of these three different discriminative mapping algorithms, other manifold learning schemes can also be used. This framework could be a reference to other related work in using deep learned feature in the age estimation problem. It can also be used as a guide for other specific problems using deep learning framework.

To analyze the performance of different classification and regression algorithms in DLA, experiments for testing performance of different regression and classification schemes are conducted in MORPH. The regression and classification algorithms include SVMs, SVR, PLS and CCA. The results are illustrated in Fig. 5 and Fig. 6. RBF kernel is used in SVMs and SVR, the parameters of the learners are adjusted based on the training data. For the parameter settings of PLS and CCA, we follow the introduction given in [21]. From the comparison of MAEs, SVR performs best in estimating age based the extracted feature (DLA). Experimental results demonstrate that SVR performs better compared to others. The performance obtained by PLS and CCA is similar to using SVMs. One possible reason is in [21], gender and ethnicity labels are also used in the training phase to learn the regression function. It is known that the aging progression across male and female, with different ethnicities are different [20, 21, 45]. In this paper, unlike [20, 21, 45], we focus on analyzing the performance of a new aging feature (DLA) instead of considering the influence due to gender and ethnicities, the label used for training is only age information. This work demonstrates that SVR outperforms PLS and CCA when only age label is provided. The results obtained by PLS and CCA are nearly the same as indicated in Fig. 5 and Fig. 6.

To compare the performance between using the feature from top layer L_5 and the extracted feature using layers from L_2 to $L5$, another experiment is conducted on MORPH where only the feature extracted from layer L_5 is used to form the deep learned aging feature. The MAE obtained from the top layer L_5 is 5.63, which is higher than the result (4.77) obtained using all the layers. But the result from the single layer L_5 is still better than the state-of-the-art. This also demonstrates the effectiveness of the proposed scheme over the state-of-the-art approaches. We think one reason why the extracted feature performs so well is that the extracted features using CNN is "supervised ". The supervised learning strategy in CNN forces the network to explore the age related information at multiple scales from images. Therefore, the learned features are more distinctive in the age information spanned manifold which can be seen in Fig. 4.

5. Conclusion

In this paper, we have proposed a new framework for face-image-based age estimation problem. A deep learning scheme based on CNN is firstly introduced to this kind of problem. We have demonstrated the effectiveness of the proposed model in predicting one's age based on the facial image. Our system is evaluated on two benchmark datasets and outperforms the state-of-the-art by a large margin in both. Manifold learning algorithm is also incorporated in the framework, showing its effectiveness in discriminative subspace learning of the deep learned age pattern. Support vector regression technique performs best among the tested regression and classification algorithms. The promising performance over the state-of-the-art works demonstrates the potential of our approach towards real world applications.

References

[1] The fg-net aging database `http://www.fgnet.rsunit.com/`.

[2] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[3] C. M. Bishop et al. Neural networks for pattern recognition. 1995.

[4] J. Bouvrie. Notes on convolutional neural networks. 2006.

[5] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[6] D. Cai, X. He, J. Han, and H.-J. Zhang. Orthogonal laplacianfaces for face recognition. *TIP*, 15(11):3608–3614, 2006.

[7] D. Cai, X. He, K. Zhou, J. Han, and H. Bao. Locality sensitive discriminant analysis. In *IJCAI*, pages 708–713, 2007.

[8] H. Chang, Y. Zhou, P. Spellman, and B. Parvin. Stacked predictive sparse coding for classification of distinct regions in tumor histopathology. In *ICCV*, pages 169–176, 2013.

[9] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *CVPR*, pages 585–592, 2011.

[10] K. Chen, S. Gong, T. Xiang, and C. C. Loy. Cumulative attribute space for age and crowd density estimation. In *CVPR*, pages 2467–2474, 2013.

[11] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.

[12] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. *PAMI*, 35(8):1915–1929, 2013.

[13] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010.

[14] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, 10(4):578–584, 2008.

[15] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *PAMI*, 29(12):2234–2240, 2007.

[16] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai. Learning from facial aging patterns for automatic age estimation. In *MM*, pages 307–316, 2006.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014.

[18] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *TIP*, 17(7):1178–1188, 2008.

[19] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer. Locally adjusted robust regression for human age estimation. *Urbana*, 51:61801, 2008.

[20] G. Guo and G. Mu. Human age estimation: What is the influence across race and gender? In *CVPRW*, pages 71–78. IEEE, 2010.

[21] G. Guo and G. Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *CVPR*, pages 657–664, 2011.

[22] G. Guo and G. Mu. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 2014.

[23] G. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *CVPR*, pages 112–119, 2009.

[24] G. Guo and X. Wang. A study on human age estimation under facial expression changes. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2547–2553, 2012.

[25] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[26] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[27] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *ICCV*, pages 2146–2153, 2009.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[29] Y. H. Kwon and N. da Vitoria Lobo. Age classification from facial images. In *CVPR*, pages 762–767, 1994.

[30] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):621–628, 2004.

[31] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *PAMI*, 24(4):442–455, 2002.

[32] B. B. Le Cun, J. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, 1990.

[33] Y. LeCun, F. J. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *CVPR*, volume 2, pages II–97, 2004.

[34] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. 2012.

[35] C. Li, Q. Liu, J. Liu, and H. Lu. Learning ordinal discriminative features for age estimation. In *CVPR*, pages 2570–2577, 2012.

[36] A. C. Little, B. C. Jones, and L. M. DeBruine. The many faces of research on face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1634–1637, 2011.

[37] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *CVPR*, volume 1, pages 387–394, 2006.

[38] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. *CVPRW*, 2014.

[39] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *FG*, pages 341–345, 2006.

[40] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.

[41] T. Serre, A. Oliva, and T. Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424–6429, 2007.

[42] L. I. Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

[43] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[44] R. D. Tobias et al. An introduction to partial least squares regression. In *20th Proc. Ann. SAS Users Group Int. Conf.*, pages 2–5, 1995.

[45] X. Wang, V. Ly, G. Lu, and C. Kambhamettu. Can we minimize the influence due to gender and race in age estimation? In *ICMLA*, volume 2, pages 309–314, 2013.

[46] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang. Ranking with uncertain labels. In *IEEE International Conference on Multimedia and Expo*, pages 96–99, 2007.

[47] S. Yan, H. Wang, X. Tang, and T. S. Huang. Learning auto-structured regressor from uncertain nonnegative labels. In *ICCV*, pages 1–8, 2007.

[48] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *PAMI*, 29(1):40–51, 2007.

[49] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. *arXiv preprint arXiv:1311.2901*, 2013.

[50] Y. Zhang and D.-Y. Yeung. Multi-task warped gaussian process for personalized age estimation. In *CVPR*, pages 2622–2629, 2010.

[51] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin. Classification of histology sections via multispectral convolutional sparse coding. In *CVPR*, 2014.