Kernel ELM and CNN based Facial Age Estimation

Furkan Gürpınar[1], Heysem Kaya[2], Hamdi Dibeklioğlu[3], and Albert Ali Salah[4]
[1]Program of Computational Science and Engineering, Boğaziçi University, Istanbul, Turkey
[2]Department of Computer Engineering, Namık Kemal University, Tekirdağ, Turkey
[3]Pattern Recognition & Bioinformatics Group, Delft University of Technology, Delft, The Netherlands
[4]Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey
furkan.gurpinar@boun.edu.tr, hkaya@nku.edu.tr,
h.dibeklioglu@tudelft.nl, salah@boun.edu.tr

Abstract

We propose a two-level system for apparent age estimation from facial images. Our system first classifies samples into overlapping age groups. Within each group, the apparent age is estimated with local regressors, whose outputs are then fused for the final estimate. We use a deformable parts model based face detector, and features from a pretrained deep convolutional network. Kernel extreme learning machines are used for classification. We evaluate our system on the ChaLearn Looking at People 2016 - Apparent Age Estimation challenge dataset, and report 0.3740 normal score on the sequestered test set.

1. Introduction

Automated age estimation from facial images is one of the most difficult challenges in face analysis. It can be very favorable in a number of real life applications such as age-based authorization systems, demographic data mining, business intelligence and video surveillance systems. The difficulty of this task originates from many reasons such as the lack of enough labeled samples to model the aging patterns of subjects, as well as uncontrolled conditions in data collection such as illumination, pose, occlusions and other environmental variables. Aging process is also known to be very subject-dependent, i.e. subjects might differ in terms of aging patterns, resulting in high variations within the samples from the same age. One of the earliest works involving age estimation from face images is conducted in early 2000s by Lanitis *et al*. [19, 18]. After the emergence of large age databases such as MORPH [32], FRGC [30] and FG-NET [1], the interest on this subject has significantly grown. In the following we provide a brief literature review on studies.

1.1. Related Work

The first steps of a human age estimation pipeline are face detection [41, 25] and facial landmark localization [43]. In this work, we chose to use the Deformable Part Model (DPM) based face detector proposed by Mathias *et al*. [25], because it finds the location of the face bounding box and gives a good alignment without the need for facial landmark localization. DPM face detector was used successfully in face-related applications such as face recognition [29], action recognition [34, 35] and age estimation [36].

A variety of feature extraction methods were applied for the task of modeling the aging pattern from facial images. For example, Active Shape Models and Active Appearance Models have been employed as features for age estimation [9, 18, 3, 24].

Histogram-based local appearance features have been a very popular choice for age estimation. These features include the Local Binary Patterns (LBP) descriptor, which encodes a local patch of the image based on the binary relations of the center pixel with its neighbors, and is widely used in age estimation [42, 22, 37, 4]. Similarly the Histogram of Oriented Gradients (HOG) descriptor has shown to be informative for age modeling [7, 22].

The LBP descriptor is modified by processing the input image with Gabor filters, resulting in the Local Gabor Binary Pattern descriptor, which has shown to be informative in age estimation [37].

Gabor filters are also employed in the calculation of Bio-Inspired Features (BIF) [33], which is consistently used for age estimation in recent years [14, 22]. BIF feature processes an image using a multi-layer feed-forward model where the first layer convolves the image with a set of Gabor filters from multiple orientations and scales, and the resulting vector is downsized with a pooling step, usually with STD or MAX operators. A simplified version of this model is used in [13], where the authors choose the number of bands and orientations manually.

Convolutional Neural Networks (CNN), originally introduced by LeCun *et al*. [20], has gained great popularity with the emergence of big datasets and computational resources. CNNs have been successfully applied to many fields of signal processing such as speech recognition [45], image retrieval [38], image classification [17] and multimodal video analysis [28].

Almost all the top-ranking participants of ChaLearn Looking at People 2015 - Apparent Age Estimation challenge used deep learning and achieved very good results [5]. Rothe *et al*. [36] won the LAP-2015 challenge by using a deep network that was trained for image classification, and fine-tuning it for the task of apparent age estimation by collecting a custom dataset of 524,230 images automatically from the Internet. Liu *et al*. [23] and Zhu *et al*. [46] also proposed to fine-tune deep networks for age estimation with augmented data, and achieved very good results.

Another important part of the age estimation pipeline is model learning, for which many different algorithms have been employed. For example, Support Vector Machine Regression (SVR) is a commonly used algorithm for this task [14, 3, 13, 42, 22].

Other learning algorithms that have been utilized in the task of age estimation include Neural Networks [18, 8], Random Forests (RF) [27, 46], projection based learners such as Partial Least Squares regression and Canonical Correlation Analysis, which are often used in combination with kernel and regularization techniques [11, 10, 12]. Ranking based methods are also commonly used for age estimation [42, 3, 44, 21].

Extreme Learning Machine (ELM) [15] is a fast and robust learning algorithm that has gained popularity over the last years with many successful applications related to face analysis such as face recognition [47, 26], smile detection [2] and emotion recognition [39, 16]. In [37], ELM is used for classification into four non-overlapping age groups and shown to yield good classification performance.

2. Methodology

In this section we describe the different parts of our age estimation pipeline, namely face alignment, feature extraction and model learning. The workflow of our proposed method is illustrated in Figure 1.

2.1. Face Alignment

We used the DPM based face detector of Mathias *et al*. [25]. The DPM face detector gives the coordinates of the bounding box (if any face is detected), as well as the detection score. Inspired by [36], we run the face detector on rotated version of the original image between -60°and 60°in 5°increments, in order to eliminate in-plane rotation. Since some of the images are rotated 90°or upside down, we also try 180°, -90°and 90°rotations. We then take the output with the maximum face score. For the cases where no face is detected, we register the whole image. Table 1 shows the number of detections on the three subsets.

Table 1: Face alignment summary

#	Train	Val	Test
Given	4113	1500	1978
Detected	4016	1462	1920

2.2. Feature Extraction

We used the deep network pretrained by Parkhi *et al*. [29] to extract CNN features from aligned images. The VGG-Face network consists of 37 layers, the final one being a 2622-dimensional softmax layer, trained for the face recognition task. We tried the performance of the final layers and found that the 33^{rd} layer, which is the first (earliest) 4096-dimensional convolution layer, was the most informative one. Therefore we used only the features from this layer in model learning. The baseline regression performances (without any grouping) of the best layers are shown in Table 2.

Table 2: Comparison of different layers of VGG-Face

Layer	Num. features	ϵ_{val}	MAE_{val}
32	25088	0.4284	4.68
33	4096	**0.4021**	**4.35**
35	4096	0.4150	4.48
37	2622	0.4066	4.38

We then normalize each feature vector by dividing it to its Euclidean norm. We have tried various normalization options prior to L_2 normalization and saw that none of them was improving the normal score, therefore we decided to use only L_2 normalization for the final system. Performance with various normalization options for the best layers are shown in Table 3.

In our experiments, we tried combinations of alternative feature normalization methods, including the sigmoid function, power normalization by 2 (i.e. setting the absolute value of each feature to its square root), min-max normalization of each feature to $[-1, 1]$ among samples, and z-normalization. For min-max and z-normalization, we learn the parameters from training folds and apply them to the test fold.

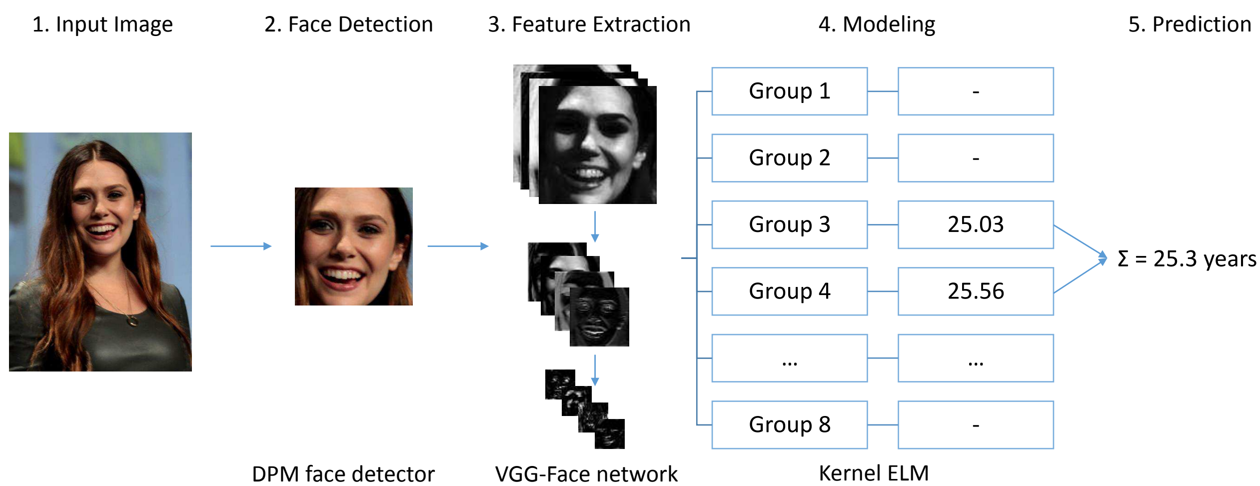

Figure 1: Pipeline of the proposed system

Table 3: Validation set performance with different normalization options

Norm. Type	Layer 33		Layer 37	
	ϵ	MAE	ϵ	MAE
No norm.	0.4487	4.91	0.4403	4.79
L_2	**0.4021**	4.35	0.4066	4.38
Pow. + L_2	0.4028	**4.32**	0.4079	4.44
Sig. + L_2	0.4152	4.49	0.4137	4.46
MM + L_2	0.4355	4.77	0.4301	4.63
Z + L_2	0.4102	4.48	0.4036	4.33
MM + Sig. + L_2	0.4861	5.46	0.4652	5.12
Z + Sig. + L_2	0.4220	4.59	0.4164	4.51
MM + Pow. + L_2	0.4565	5.01	0.4438	4.88
Z + Pow. + L_2	0.4083	4.43	0.4078	4.37

It should be noted that the final classification layer also yielded a very informative feature, but it did not improve the overall accuracy when fused with the 33^{rd} layer output. Considering that this deep network is trained for face recognition, relatively good performance of the last layer also indicates the informativeness of age in discriminating identity from facial image.

2.3. Modeling

In this section we explain the two parts of our age modeling system; classification into overlapping age groups and regression among each group. We then fuse the decisions of each relevant group to obtain the final estimation.

2.3.1 Kernel ELM

In both stages of our model learning pipeline, we used kernel Extreme Learning Machines (ELM) [15] due to the learning speed and accuracy of the algorithm. In the following paragraphs, we briefly explain the learning strategy of ELMs.

ELM proposes a simple and robust learning algorithm for single-hidden layer feedforward networks. The input layer's bias and weights are initialized randomly to obtain the output of the second (hidden) layer. The bias and weights of the second layer are calculated by a simple generalized inverse operation of the hidden layer output matrix.

ELM tries to find the mapping between the hidden node output matrix $\mathbf{H} \in \mathbb{R}^{N \times h}$ and the label vector $\mathbf{T} \in \mathbb{R}^{N \times 1}$ where N and h denote the number of samples and the hidden neurons, respectively. The set of output weights $\beta \in \mathbb{R}^{h \times 1}$ is calculated by the least squares solution of the set of linear equations $\mathbf{H}\beta = \mathbf{T}$, as:

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{1}$$

where $\mathbf{H}^{\dagger}$ denotes the Moore-Penrose generalized inverse [31] that minimizes the L_2 norms of $||\mathbf{H}\beta - \mathbf{T}||$ and $||\beta||$ simultaneously.

To increase the robustness and the generalization capability of ELM, a regularization coefficient $\mathbf{C}$ is included in the optimization procedure. Therefore, given a kernel $\mathbf{K}$, the set of weights is learned as follows:

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{K}\right)^{-1}\mathbf{T}. \tag{2}$$

We use the radial basis function (RBF) in order to calculate kernel $\mathbf{K}$ from the original features because in our preliminary experiments we found that its performance is was superior compared to its alternatives such as linear and polynomial kernels. In Table 4, we provide the performances of different kernel types for global regression with the 33^{rd} layer output.

Table 4: Validation performance with different kernel types

Linear		RBF		Poly	
ϵ	MAE	ϵ	MAE	ϵ	MAE
0.49	5.26	0.42	4.35	0.45	4.48

As explained in Section 2.2, we normalize the feature vectors by their L_2 norm before comparing them with RBF, so we obtain a normalized kernel representation that enables us to model the aging pattern in a more generalized way.

2.3.2 Classification

For each age group $(x \mid a_1 \leq x \leq a_2)$, we learn a binary classification model by labeling the training set based on membership to the group and optimizing the parameters of kernel ELM classifier by random 3-fold cross validation within the training set. Therefore, we run G binary classifiers for each sample in the validation set to determine which groups it belongs to, in order to get estimations only from those groups. We tried to optimize the models based on binary classification accuracy as well as recall, and we found that optimizing based on accuracy yielded slightly better results, therefore we chose to optimize based on accuracy. The final performance of our system on individual groups, as well as the selected classification parameters, are summarized in Table 5.

2.3.3 Regression

We learn a regression model for each age group simply by choosing the positive samples from the training set and optimizing the parameters of kernel ELM regressor by random 3-fold cross validation on this subset. For each sample in the validation set, we first run the binary classifiers for each group, and for the ones with positive response, we apply the corresponding local regression models to get their estimations. Therefore, we obtain g estimations per instance, where $0 \leq g \leq G$. For the samples which are not assigned to any group (this was the case for 11% of the validation and 15% of the test set), we learn a backup model from all the samples in the training set, that is the global regression model which is equivalent to using the age group $[0, \infty]$ (see Table 5).

3. Experiments

In this section we describe the challenge corpus, experimental setup and the evaluation protocol, and we present the performance of our classification and regression methods with example images from the validation and test sets.

3.1. LAP-2016 Dataset

ChaLearn Looking at People 2016 - Apparent Age Estimation challenge dataset [6] consists of 7,591 face images collectively labeled by multiple human annotators, therefore the mean μ and the standard deviation σ is provided for each sample. The dataset is split into 4113 training, 1500 validation and 1978 testing samples, where the testing set labels are sequestered. The three subsets have a similar age distribution. Table 1 presents the number of samples where the DPM face detector was able to detect a face.

3.2. Evaluation Criteria

Mean Absolute Error (MAE): A standard way of measuring the accuracy of a regressor is to average the absolute deviation of each sample's label from its estimated value. More formally, MAE of a given dataset is calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |x_i - \hat{x}_i|, \tag{3}$$

where x_i is the true label i.e. the average of apparent age annotations for sample i, $\hat{x}_i$ is the predicted value, and N is the number of testing samples.

Normal Score (ϵ): Since the LAP-2016 dataset is labeled by multiple annotators, the performance of an age estimation system might be more accurately measured by taking into account the variance of the annotations for each sample. Therefore the ϵ-score is calculated by fitting a normal distribution with mean μ and standard deviation σ of the annotations for each sample:

$$\epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

Thus, the average ϵ-score for a dataset can change between 1 (worst case) and 0 (best case).

3.3. Implementation details

The system is implemented in MATLAB. Face detection takes around 2 seconds per image and rotation angle. Feature extraction from VGG-Face with MatConvNet library [40] takes around 1 second per image. For classification and regression, we optimize the kernel parameter γ and the regularization coefficient C with a grid search where both parameters are searched in the exponential set $2^{[-2,-1,\dots,6]}$. Training the whole system takes 12 minutes and obtaining the estimation takes around 2 seconds per test image.

Table 5: Classification accuracy, recall and regression performance on validation set with different age groups. N denotes the number of samples

Group	N_{tr}	N_{val}	Acc.	Rec.	ϵ	MAE
0-15	860	152	0.96	0.78	0.45	2.46
10-25	2366	436	0.84	0.65	0.31	2.90
15-30	3686	662	0.84	0.83	0.31	3.19
20-35	4072	705	0.81	0.86	0.33	3.52
30-40	1764	311	0.81	0.35	0.34	3.82
35-50	1568	288	0.85	0.45	0.34	4.26
45-60	976	184	0.91	0.48	0.28	3.87
55-∞	554	106	0.96	0.57	0.28	4.36
0-∞	8032	1462	-	-	0.40	4.35
Overall	8032	1462	-	-	0.33	3.85

3.4. Results

In this section we present the results of our classification and regression systems. In Table 5, we summarize the classification accuracy and recall for the 8 overlapping age groups we used. The 9^{th} row is the performance of the backup system, and the final row is the performance of the whole system on the validation set of LAP-2016 dataset.

Table 5 shows that the ensemble of local regressors yield smaller MAE for younger age groups. As the age progresses, within-group variance increases with it, making the apparent age estimation task harder. Finally, since younger subjects are usually annotated with less variance, the ϵ-score behaves almost inversely to MAE score, as the ϵ-score is more tolerant for the errors in the older subjects.

We display the estimation results on samples from the validation set in Figures 2 and 3. Figure 2 shows the invariance of CNN features to common difficulties such as blur, pose and occlusions. In Figure 3, we show the examples where our age estimation system fails due to many possible reasons such as mis-detection of the face, alignment errors or simply due to lack of enough samples to model the aging pattern, especially for older subjects.

Our system achieves 0.33 ϵ-score in the development phase, and 0.37 ϵ-score in the test phase of the challenge. The final results of the challenge is displayed in Table 6.

4. Conclusions and Future Work

In this paper, we propose an apparent age estimation system with the use of deep learning and a fast and robust age modeling algorithm. We show that the performance of local regressors are better than the global regressor for almost all groups. However, we give equal weight to each group a sample is assigned to, whereas weighing the decisions with a membership score can result in more accurate estimation.

Table 6: ChaLearn Looking At People 2016 Apparent Age Estimation challenge final results

Position	Team	Test error
1	OrangeLabs	0.2411
2	palm_seu	0.3214
3	cmp+ETH	0.3361
4	WYU_CVL	0.3405
5	ITU_SiMiT	0.3668
6	Bogazici (Ours)	0.3740
7	MIPAL_SNU	0.4569
8	DeepAge	0.4573

CNNs are robust to common difficulties in image processing such as pose and illumination differences as well as occlusions. Therefore our system works with a very coarse alignment system, however we believe that obtaining a finer alignment with the help of a landmark detection system will further improve the estimation accuracy.

We make use of transfer learning by using the features from a deep network that is trained on a face recognition task and directly employing them in age estimation. However, literature has shown that fine-tuning a deep network for the age estimation task increases the explanatory power of its features, therefore in the future we aim to train a convolutional neural network for real and apparent age estimation tasks.

Acknowledgment

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 114E481.

References

[1] The FG-NET aging database, November 2002. Available at `http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm`.

[2] L. An, S. Yang, and B. Bhanu. Efficient smile detection by extreme learning machine. *Neurocomputing*, 149:354–363, 2015.

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

[4] H. Dibeklioğlu, F. Alnajar, A. Ali Salah, and T. Gevers. Combining facial dynamics with appearance for age estimation. *IEEE Trans. Image Processing*, 24(6):1928–1943, 2015.

[5] S. Escalera, J. Fabian, P. Pardo, X. Baro, J. Gonzalez, H. Escalante, D. Misevic, U. Steiner, and I. Guyon. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In *ICCVW, 2015*, pages 1–9.

[6] S. Escalera, M. Torres, B. Martnez, X. Bar, H. J. Escalante, I. Guyon, G. Tzimiropoulos, C. Corneanu, M. Oliu, M. A.

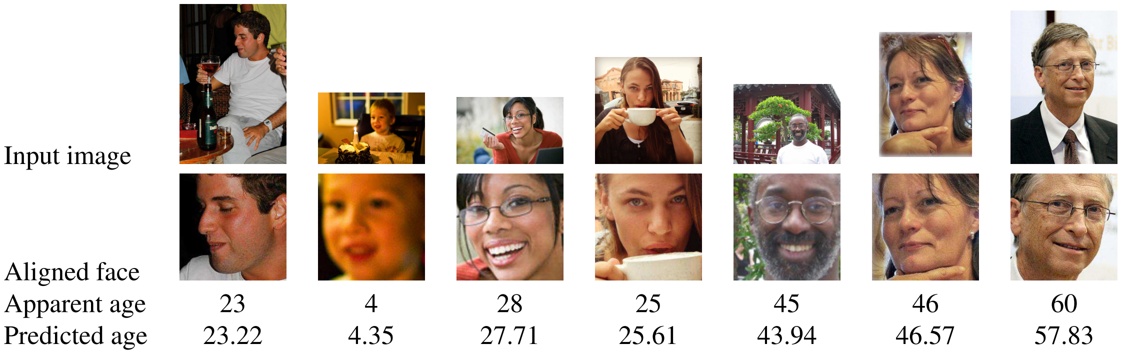

Figure 2: Examples from the validation set with good estimations

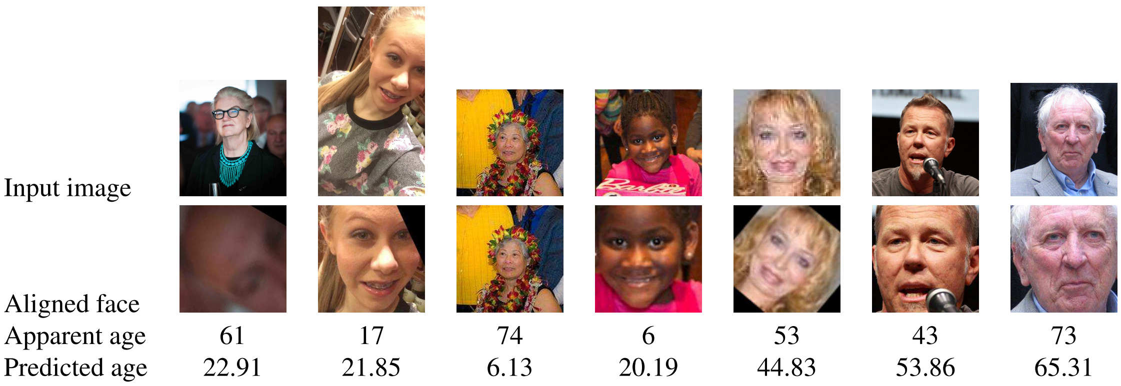

Figure 3: Examples from the validation set with bad estimations

Bagheri, and M. Valstar. Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016. In *ChaLearn Looking at People and Faces of the World, CVPRW*, 2016.

[7] C. Fernández, I. Huerta, and A. Prati. A comparative evaluation of regression learning algorithms for facial age estimation. In *Face and Facial Expression Recognition from Real World Videos*, pages 133–144. 2015.

[8] X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. *TPAMI*, 35(10):2401–2412, 2013.

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

[10] G. Guo and G. Mu. Joint estimation of age, gender and ethnicity: Cca vs. pls. In *FG, 2013*, pages 1–6.

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

[12] G. Guo and G. Mu. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 32(10):761–770, 2014.

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

[14] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *ICB, 2013*, pages 1–8.

[15] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. B*, 42(2):513–529, 2012.

[16] H. Kaya, F. Gürpinar, S. Afshar, and A. A. Salah. Contrasting and combining least squares based learners for emotion recognition in the wild. In *ACM ICMI*, pages 459–466, 2015.

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

[18] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern. B*, 34(1):621–628, 2004.

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

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[21] S. Li, S. Shan, and X. Chen. Relative forest for attribute prediction. In *ACCV, 2012*, pages 316–327. 2012.

[22] K.-H. Liu, S. Yan, and C.-C. J. Kuo. Age estimation via grouping and decision fusion. *TIFS*, 10(11):2408–2423, 2015.

[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 *ICCVW, 2015*, pages 16–24.

[24] K. Luu, K. Ricanek Jr, T. D. Bui, and C. Y. Suen. Age estimation using active appearance models and support vector machine regression. In *BTAS, 2009*, pages 1–5.

[25] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face detection without bells and whistles. In *ECCV, 2014*, pages 720–735.

[26] A. A. Mohammed, R. Minhas, Q. J. Wu, and M. A. Sid-Ahmed. Human face recognition based on multidimensional pca and extreme learning machine. *Pattern Recognition*, 44(10):2588–2597, 2011.

[27] A. Montillo and H. Ling. Age regression from faces using random forests. In *ICIP, 2009*, pages 2465–2468.

[28] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *ICML, 2011*, pages 689–696.

[29] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, page 6, 2015.

[30] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR, 2005*, pages 947–954.

[31] C. R. Rao and S. K. Mitra. *Generalized inverse of matrices and its applications*, volume 7. Wiley New York, 1971.

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

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

[34] A. Rosenfeld and S. Ullman. Face-space action recognition by face-object interactions. *arXiv preprint arXiv:1601.04293*, 2016.

[35] A. Rosenfeld and S. Ullman. Hand-object interaction and precise localization in transitive action recognition. *arXiv preprint arXiv:1511.03814*, 2016.

[36] R. Rothe, R. Timofte, and L. Gool. Dex: Deep expectation of apparent age from a single image. In *ICCVW, 2015*, pages 10–15.

[37] P.-K. Sai, J.-G. Wang, and E.-K. Teoh. Facial age range estimation with extreme learning machines. *Neurocomputing*, 149:364–372, 2015.

[38] A. Torralba, R. Fergus, and Y. Weiss. Small codes and large image databases for recognition. In *CVPR, 2008*, pages 1–8.

[39] P. Utama, H. Ajie, et al. A framework of human emotion recognition using extreme learning machine. In *ICAICTA, 2014*, pages 315–320. IEEE, 2014.

[40] A. Vedaldi and K. Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 689–692. ACM, 2015.

[41] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[42] R. Weng, J. Lu, G. Yang, and Y.-P. Tan. Multi-feature ordinal ranking for facial age estimation. In *FG, 2013*, pages 1–6.

[43] X. Xiong and F. De la Torre. Supervised Descent Method and Its Application to Face Alignment. In *CVPR, 2013*, pages 532–539, 2013.

[44] P. Yang, L. Zhong, and D. Metaxas. Ranking model for facial age estimation. In *ICPR, 2010*, pages 3404–3407. IEEE, 2010.

[45] D. Yu, F. Seide, G. Li, and L. Deng. Exploiting sparseness in deep neural networks for large vocabulary speech recognition. In *ICASSP, 2012*, pages 4409–4412.

[46] Y. Zhu, Y. Li, G. Mu, and G. Guo. A study on apparent age estimation. In *ICCVW, 2015*, pages 25–31.

[47] W. Zong and G.-B. Huang. Face recognition based on extreme learning machine. *Neurocomputing*, 74(16):2541–2551, 2011.