

Automatic Age and Gender Estimation using Deep Learning and Extreme Learning Machine

Anto A Micheal^a, R Shankar^b

^a Associate Professor, Department of Computer Science and Engineering, Teegala Krishna Reddy Engineering College, Hyderabad.

^b Professor, Department of Electronics and Communications Engineering, Teegala Krishna Reddy Engineering College, Hyderabad

Article History: *Do not touch during review process(yyyy)*

Abstract: Age and gender classification has become applicable to an extending measure of applications, particularly resulting to the ascent of social platforms and social media. Regardless, execution of existing strategies on real-world images is still fundamentally missing, especially when considered the immense bounced in execution starting late reported for the related task of face acknowledgment. In this paper we exhibit that by learning representations through the use of significant Convolutional Neural Network (CNN) and Extreme Learning Machine (ELM). CNN is used to extract the features from the input images while ELM classifies the intermediate results. We experiment our architecture on the recent Adience benchmark for age and gender estimation and demonstrate it to radically outflank current state-of-the-art methods. Experimental results show that our architecture outperforms other studies by exhibiting significant performance improvement in terms of accuracy and efficiency.

Keywords: Age Estimation, Gender Recognition, Convolutional Neural Network (CNN), Extreme Learning Machine (ELM)

1. Introduction

Age and gender assume essential parts in social between activities. Dialects hold distinctive greetings and grammar rules for men or women, and frequently diverse vocabularies are utilized while tending to senior citizens compared to youngsters [1]. In spite of the essential parts these characteristics play in our everyday lives, the capacity to consequently assess them precisely and dependably from face image is still a long way from addressing the requirements of business applications [5]. This is especially puzzling while considering late claims to super-human capacities in the related errand of face recognition. (e.g. [48]). Past ways to deal with assessing or ordering these properties from face images have depended on contrasts in facial feature dimensions [29] or "customized" face descriptors (e.g., [10, 15, 32]). Most have utilized characterization plans composed especially for age or gender orientation estimation undertakings, including [4] and others. The past strategies were intended to handle the numerous difficulties of unconstrained imaging conditions [10]. In addition, the machine learning strategies utilized by these frameworks did not completely abuse the huge quantities of image cases and information accessible through the Internet keeping in mind the end goal to enhance characterization capacities.

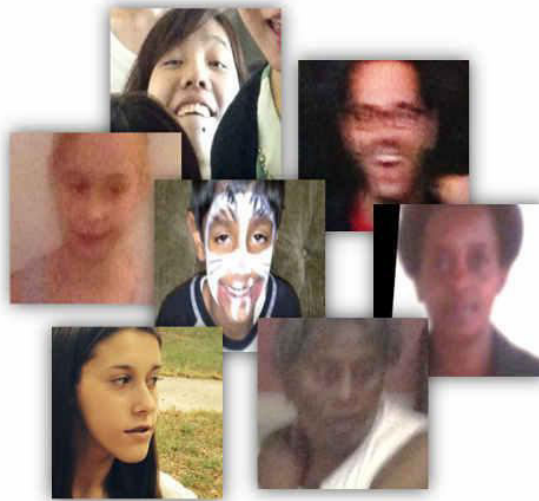
In this paper, the endeavor is to close the gap between automatic face recognition abilities and those of age and gender classification techniques. To this end, we take after the fruitful sample set around late face recognition frameworks: Face recognition systems portrayed in the most recent couple of years have demonstrated that gigantic advancement can be made by the utilization of profound convolutional neural networks (CNN) [31]. To the best of our knowledge, SVM, Naive Bayes [7], and Extreme Learning Machine (ELM) [8] are three important classification algorithms at present while ELM has been proved to be an efficient and fast classification algorithm because of its good generalization performance, fast training speed, and little human intervene [9]. When ELM is combined with CNN it gives a good performance [10]. We show comparative results, composed by considering the somewhat constrained accessibility of precise age and gender classification names in existing face information sets. The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 discusses architecture of CNN–ELM model. The experiments and results are illustrated in Section 4 and 5. Finally, it is concluded in Section 6.

2. Related Work

2.1. Age Classification

The issue of consequently extricating age related traits from facial images has got expanding consideration as of late and numerous strategies have been put forth. A point by point overview of such strategies is found in [11] and, all the more as of late, in [21]. We take note of that regardless of our attention here on age group characterization as opposed to exact age estimation (i.e., age regression), the study incorporates strategies intended for either undertaking. Early techniques for age estimation depend on ascertaining proportions between various estimations of facial features [29]. When facial features (e.g. eyes, nose, mouth, jaw, and so forth.) are confined, their sizes and separations are measured, proportions between them are ascertained and utilized for arranging the face into various age classifications as indicated by hand-made principles [12]. All the more as of late, [41] utilizations a comparative way to deal with model age movement in subjects less than 18 years of age [22]. As those techniques require precise restriction of facial elements, testing issues are independent from anyone else, they are unacceptable for in-the-wild images which one might hope to discover on social platform.

Figure.1 Faces from the Adience benchmark for age and gender classification [10]



On a substitute calling are strategies that address the developing procedure as a subspace [16] or a complex [19]. An impediment of those systems is that they require information about the image to be close frontal and all that much balanced. These systems in like manner present test comes to fruition just on constrained data sets of close frontal images (e.g UIUC-IFP-Y [12, 19], FG-NET [30] and MORPH [43]). Again, accordingly, such strategies are ill-suited for unconstrained images. Not exactly the same as those depicted above are methods that usage adjacent components for identifying with face images. In [55] Gaussian Mixture Models (GMM) [13] were used to the scattering of facial patches. In [54] GMM were used again to speak to the scattering of close-by facial estimations, however effective descriptors were used as opposed to pixel patches. Finally, instead of GMM, Hidden-Markov-Model, super-vectors [40] was used as a piece of [56] face patch transports.

A different option for the neighbourhood image force patches are vigorous image descriptors: Gabor image descriptors [32] were utilized as a part of [15] alongside a Fuzzy-LDA classifier which considers a face images as fitting in with more than one age class. In [20] a blend of Biologically-Inspired Features (BIF) [44] and different complex learning techniques were utilized for age estimation. Gabor [32] and nearby twofold examples (LBP) [1] components were utilized as a part of [7] alongside a various levelled age classifier made out of Support Vector Machines (SVM) [9] to order the info image to an age-class took after by a bolster vector relapse [52] to appraise an exact age. At last, [4] proposed enhanced forms of important part investigation [3] and locally safeguarding projections [36]. Those techniques are utilized for separation learning and dimensionality diminishment, separately, with Active Appearance Models [8] as an image highlight. These techniques have demonstrated successful on little and/or obliged benchmarks for age estimation [26]. As far as anyone is concerned, the best performing techniques were shown on the Group Photos benchmark [14]. In [10] best in class execution on this benchmark was exhibited by utilizing LBP descriptor varieties [53] and a dropout-SVM classifier. We demonstrate our proposed technique to beat the outcomes they give an account of all more difficult Adience benchmark Fig. 1, intended for the same errand.

2.2. Gender Classification

A point by point study of gender classification arrangement techniques can be found in [34] and all the more as of late in [42]. Here we rapidly review significant strategies. One of the early techniques for gender classification characterization [17] utilized a neural network system prepared on a little arrangement of close frontal face images. In [37] the consolidated 3D structure of the head (acquired utilizing a laser scanner) and image intensities were utilized for grouping gender classification [45]. SVM classifiers were utilized by [35], connected specifically to image intensities. Instead of utilizing SVM [2], utilized AdaBoost for the same reason, here once more, connected to image intensities. At long last, perspective invariant age and gender classification characterization was presented by [49]. All the more as of late, [51] utilized the Weber's Local composition Descriptor [6] for gender classification acknowledgment, exhibiting close immaculate execution on the FERET benchmark [39]. In [38], power, shape and surface elements were utilized with shared data, again getting close immaculate results on the FERET benchmark.

A large portion of the strategies talked about the above utilized FERET benchmark [39] both to build up the proposed frameworks and to assess exhibitions. FERET images were taken under profoundly controlled condition and are along these lines considerably less difficult than in-the-wild face images. In addition, the outcomes got on this benchmark propose that it is soaked and not trying for present day strategies. It arrives fore hard to appraise the genuine relative advantage of these methods. As an outcome, [46] probed the prominent Labelled Faces in the Wild (LFW) [25] benchmark, basically utilized for face acknowledgment. Their technique is a blend of LBP components with an AdaBoost classifier. Likewise with age estimation, here as well, we concentrate on the Adience set which contains images more difficult than those gave by LFW, reporting execution utilizing a heartier framework, intended to better adventure data from monstrous illustration preparing sets.

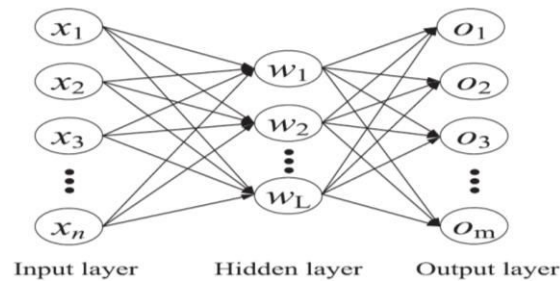
2.3. Convolutional Neural Networks

One of the primary utilizations of convolutional neural networks (CNN) is maybe the LeNet-5 system depicted by [31] for optical character acknowledgment. Contrasted with current profound CNN, their system was generally humble because of the restricted computational assets of the time and the algorithmic difficulties of preparing greater systems. In spite of the fact that much potential laid in more profound CNN designs (systems with more neuron layers), just as of late have they got to be predominant, after the emotional increment in both the computational force, the measure of preparing information promptly accessible on the Internet, and the improvement of more viable techniques for preparing such complex models. One later and remarkable case is the utilization of profound CNN for image classification based on the testing Image net benchmark [28]. Profound CNN have moreover been effectively connected to applications including human posture estimation [50], face parsing [33], facial key point identification [47], discourse acknowledgment [18] and activity characterization [27].

2.4. Extreme Learning Machine Model

ELM was first proposed by Huang et al. [58, 60, 61] which was used for the single-hidden-layer feed forward neural networks (SLFNs). The input weights and hidden layer biases are randomly assigned at first, and then the training datasets to determine the output weights that are combined. The basic structure of ELM is shown in Figure 2.

Figure.2. The Structure of ELM

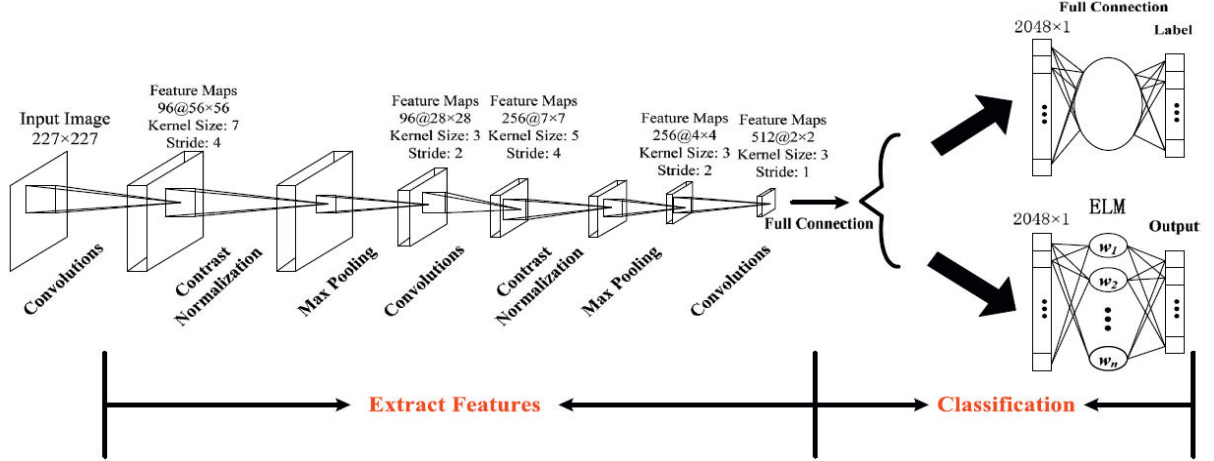


ELM is not only widely used to process binary classification [62–65], but also used for multi-classification due to its good properties. As CNNs show excellent performance on extracting feature from the input images, which can reflect the important character attributes of the input images. Therefore, we can integrate the advantages of CNNs and ELM based on the analysis above, which means CNNs extract features from the input images while ELM classify the input feature vectors.

3. Methodology

Fig. 3 shows the architecture of our CNN-ELM. It can be seen from the figure that our network includes two stages, feature extraction and classification. The stage of feature extraction contains the convolutional layer, contrast normalization layer, and max pooling layer. The first convolutional layer consists of 96 filters, and the size of its feature map is 56×56 while its kernel size is 7 and the stride of the sliding window is 4. A single convolution layer is implemented after the two stages, and a full connection layer converts the feature maps into 1-D vectors which is beneficial to the classification. Finally, the ELM structure is combined with the designed CNN model, and this architecture is used to classify the age and gender tasks.

Figure.3. The architecture of Age and Gender detection using CNN+ELM Model



3.1. Convolutional Layer

In the convolutional layer, convolutions are performed between the previous layer and a series of filters, extract features from the input feature maps [66, 67]. After that, the outputs of the convolutions will add an additive bias and an element-wise non-linear activation function is applied on the front results. ReLU function is used as the nonlinear function in the experiment.

In general, η_{ij}^{mn} denotes the value of an unit at position (m, n) in the j^{th} feature map in the i^{th} layer and it can be expressed as Eq. (1) :

$$\eta_{ij}^{mn} = \sigma \left(b_{ij} + \sum_{\delta} \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ij\delta}^{pq} \eta_{(i-1)\delta}^{(m+p)(n+q)} \right) \quad (1)$$

Where b_{ij} represents the bias of this feature map while δ indexes over the set of the feature maps in the $(i-1)^{\text{th}}$ layer which are connected to this convolutional layer. $w_{ij\delta}^{pq}$ denotes the value at the position (p,q) of the kernel which is connected to the k^{th} feature map and the height and width of the filter kernel are P_i and Q_i .

3.2. Contrast Normalization Layer

The goal of the local contrast normalization layer is not only to enhance the local competitions between one neuron and its neighbours, but also to force features of different feature maps in the same spatial location to be computed [67]. In order to achieve the target, two normalization operations, i.e., subtractive and divisive, are performed.

3.3. Max Pooling Layer

The purpose of pooling strategy is to transform the joint feature representation into a novel, more useful one which keeps crucial information while discards irrelevant details. Each feature map in the sub sampling layer is getting by max pooling operations which are carried out on the corresponding feature map in convolutional layers [62]. Eq. (2) is the value of a unit at position (m,n) in the j^{th} feature map in the i^{th} layer or sub sampling layer after max pooling operation:

$$\eta_{ij}^{mn} = \max \left\{ \eta_{(i-1)j}^{mn}, \eta_{(i-1)j}^{(m+1)(n+1)}, \dots, \eta_{(i-1)j}^{(m+P_i)(n+Q_i)} \right\} \quad (2)$$

The max pooling operation generates position invariance over larger local regions and down samples the input feature maps. In this time, the numbers of feature maps in the sub sampling layer are 96 while the size of the filter is 3 and the stride of the sliding window is 2. The aim of max pooling action is to detect the maximum response of the generated feature maps while reduces the resolution of the feature map. Moreover, the pooling operation also offers built-in invariance to small shifts and distortions.

3.4. ELM Classification Layer

After the convolution and sub sampling operations, ELM is used to classify the 1-D vectors which are converted from feature maps. The ELM updates the output weights while input weights and hidden-layer biases are randomly set, thus we will randomly generate the input parameters and calculate the output weights during the training stage [63]. The whole process without iteration operation improves the neural network generalization ability. Fig. 3 shows the output (containing 2048 × 1 dimensionality) of full-connection layer is the input of ELM while the numbers of hidden nodes are variables. The connection between ELM and convolutional network is a critical process and we can see from Fig. 3 that our input of ELM is the output of the full connection layer whose preceding layer is a convolutional layer. Forward-propagation and back-propagation operations are the principal parts in the architecture.

3.5. Process of our CNN–ELM

The steps are summarized as follows:

Step 1: Tune the parameters of CNN during the training stage when the connection between convolutional layers and output labels is full connection layers.

Step 2: Compute the hidden layer weights and cache the intermediate β matrices, meanwhile verify the accuracy of fine-tuned network.

Step 3: Stop the training process and calculate the average of β .

Step 4: Classify the unknown dataset using the architecture.

In order to fine tune the network, the structure is trained for more than 10K iterations. This process is performed to tune the parameters of CNN and makes it own the ability of extracting discriminative features.

3.6. Training Stage using Hybrid Structure

The training stage not only tunes the parameters of convolutional layer, but also achieves the corresponding hidden layer weights of ELM. The feed-forward process of the architecture is as same as a plain CNN for every 1000 iterations, ELM layers, instead of full connection layers, will be invoked and corresponding hidden layer weights are calculated [66]. At the same time, intermediate results β matrices are stored in the memory for final average results. When ELM classifier works and the whole iterations continue, the system will adopt stochastic gradient descent to tune the relevant parameters of the entire convolutional networks. During process of back propagation, the operations between convolutional layer and sub sampling layer or sub sampling layer and convolutional layer are as same as a single convolution neural network. After that, the local gradient is computed in the full connection layer. Compared with a plain CNN, the proposed architecture transforms the feature maps into 1-D vectors in the process of forward propagation, so it is just needed to transform the local gradient in the input layer of ELM to convolutional layer.

3.7. Classification Process

The structure is fine tuned and its accuracy meeting is verified. We classify the unknown subjects into different age or gender categories. The information is extracted from input dataset to hidden layers, and then classified as corresponding output.

The steps are as follows:

Step 1: Extract the features with convolutional layers from the unknown subjects.

Step 2: Classify the features using our fine-tuning structure.

We have found that little misalignments in the Adience images, brought on by the numerous difficulties of these images (impediments, movement obscure, and so forth.) can noticeably affect the nature of our outcomes. This second, over-testing strategy is intended to adjust for these misalignments, bypassing the requirement for

enhancing arrangement quality, yet rather specifically bolstering the system with different interpreted adaptations of the same face.

4. Experiment

The Adience benchmark: the precision of our CNN plan utilizing the as of late discharged Adience benchmark [10], intended for age and gender classification. The Adience image set comprises of images consequently transferred to Flickr from PDA gadgets. Since these images were transferred without former manual sifting, as is ordinarily the case on media site pages (e.g., images from the LFW gathering [25]) or social network sites (the Group Image set [14]), the conditions in these images are exceedingly unconstrained, reflecting a significant number of this present difficulties of confronts showing up in networking images. Adience images along these lines catch compelling varieties in head posture, lightning conditions quality, and more, which mean that the photos are taken without careful preparation or posing.

The whole Adience image set gathering incorporates around 26K images of 2,284 subjects. Table 1 records the breakdown of the accumulation into the distinctive age classifications. Testing for both age and gender is performed utilizing a standard five-fold, subject-selective cross-approval convention, characterized in [10]. We utilize the in-plane adjusted adaptation of the countenances, initially utilized as a part of [10]. These images are utilized as opposed to more up to date arrangement procedures so as to highlight the execution pick up ascribed to the system design, as opposed to better pre-processing. The test time with same system design and utilized for all test folds of the benchmark and indeed, for both gender and age estimation assignments. This is performed with a specific end goal to guarantee the legitimacy of our outcomes crosswise over folds, additionally to show the sweeping statement of the system plan proposed here; the same engineering performs well crosswise over various, related issues. We contrast beforehand reported results with the outcomes processed by our system. Our outcomes incorporate two techniques for testing: center crop and over-sampling.

Table.1. The Adience Faces Benchmark

Gender /Years	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19587

5. Results

Table 2 shows our outcomes for gender and age classification separately and Fig. 8 shows the graphical representation of the accuracy. Table 3 further gives a confusion matrix to our multi-class age grouping results. For age arrangement, we measure and look at both the exactness when the calculation gives the precise age-bunch order and when the algorithm is off by one nearby age-bunch (i.e., the subject fits in with the gathering instantly more seasoned or quickly more youthful than the anticipated gathering). This tails other people who have done as such before, and reflecting the instability natural to the errand – facial components frequently change next to no between most seasoned countenances in one age class and the most youthful appearances of the consequent class.

Both tables contrast execution and the strategies depicted in [10]. Table 2 additionally gives a correlation [23] which utilized the same gender classification pipeline of [10] connected to more compelling arrangement of the countenances; faces in their tests were artificially adjusted to show up confronting forward. Clearly, the proposed strategy beats the reported cutting edge on both assignments with impressive considerable gaps. Likewise, obvious is the commitment of the over-examining approach, which gives an extra execution support over the first system. This suggests better arrangement (e.g., frontalization [22, 23]) might give an extra support in execution.

The result of the age and gender estimated using the Conventional Neural Network (CNN) and ELM is shown in Fig. 4 and Fig. 5 respectively. We give a couple of samples of both gender and age misclassifications in Fig. 6 and Fig. 7, separately. These demonstrate that a large number of the errors made by our framework are because of a great degree testing seeing states of a percentage of the Adience benchmark images. Most outstanding are mix-ups brought on by obscure or low determination and impediments (especially from substantial cosmetics). Gender estimation confuses likewise habitually happen for images of infants or exceptionally youthful kids where evident gender traits are not yet noticeable.

Table.2. Gender Estimation Results on the Adience Benchmark

Method	Accuracy
Support Vector Machine [10]	77.8±1.3
3D face shape estimation [23]	79.3±0.0
CNN using single crop	85.9±1.4
CNN using over-fitting	86.8±1.4
Proposed CNN-ELM	90.2±1.2

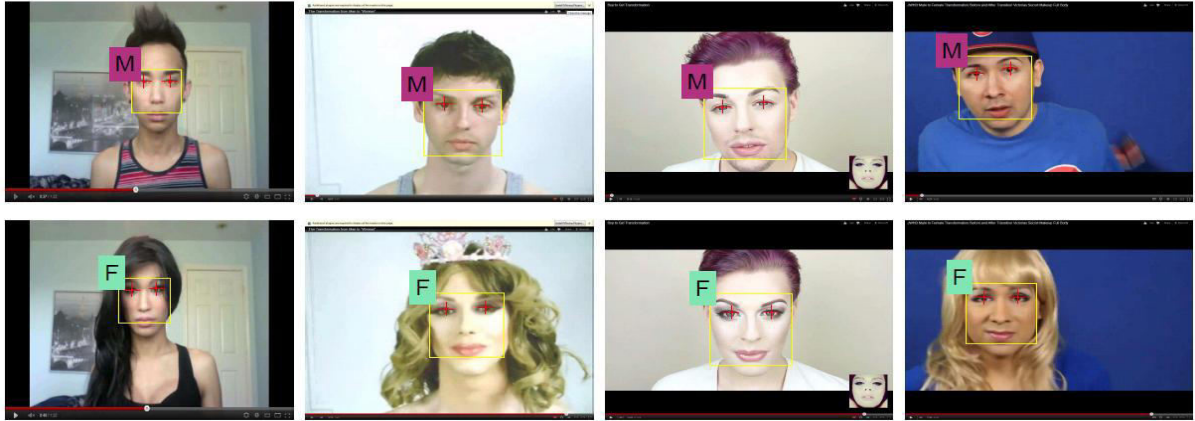
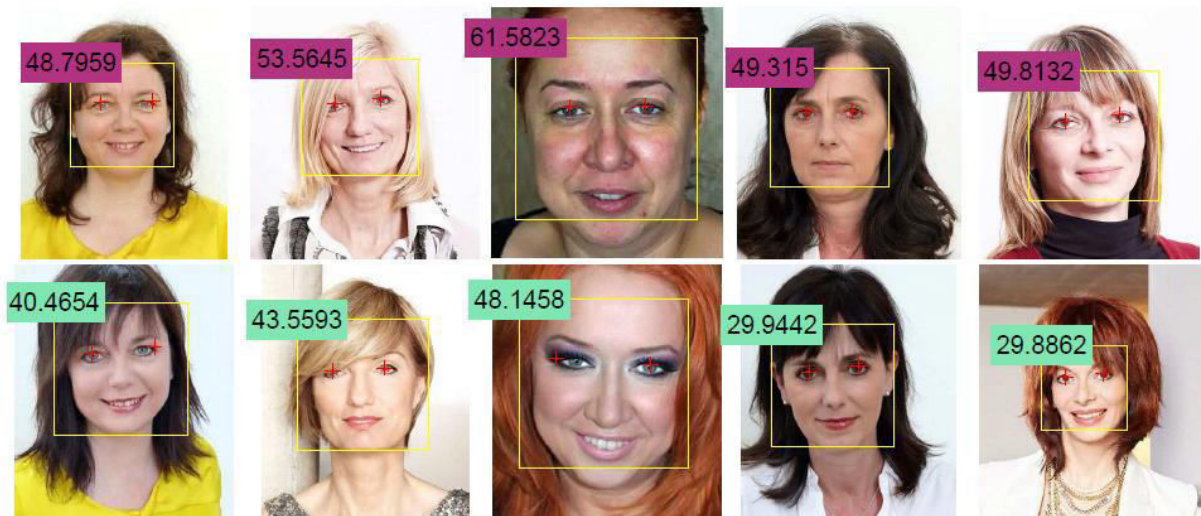
Figure.4. The results of Automatic gender recognition using the proposed approach**Figure.5.** The results of Automatic age estimation using the proposed approach**Figure.6.** Gender misclassifications. Top row: Female subjects mistakenly classified as males. Bottom row: Male subjects mistakenly classified as females

Figure.7. Age misclassifications. Top row: Older subjects mistakenly classified as younger. Bottom row: Younger subjects mistakenly classified as older



Figure.8. Accuracy comparison for Gender Estimation Methods

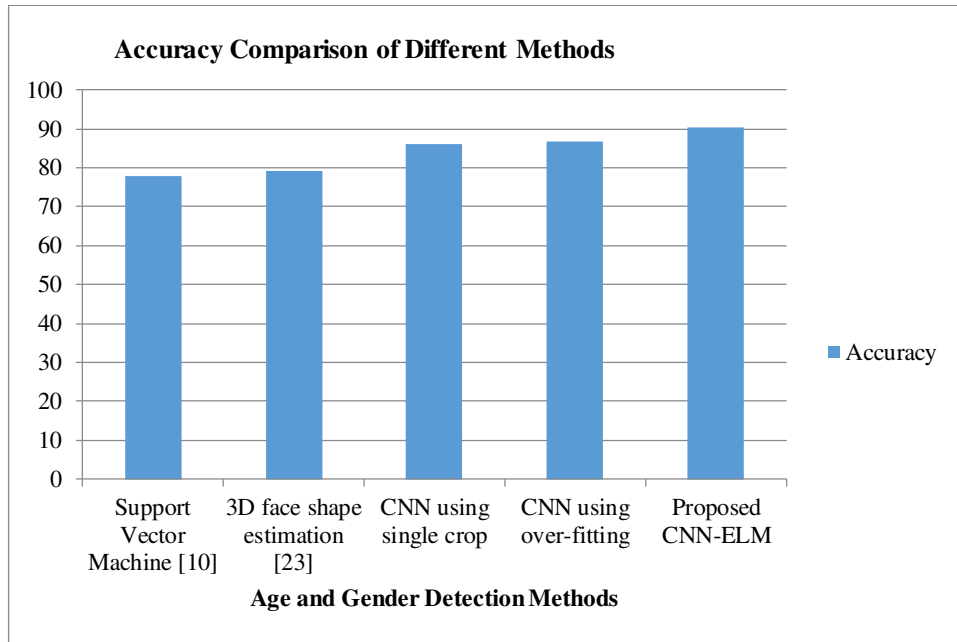


Table.3. Age Estimation Confusion Matrix on the Adience Benchmark

Age Range in Years	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-
0-2	0.753	0.147	0.028	0.006	0.005	0.008	0.007	0.009
4-6	0.256	0.652	0.166	0.023	0.010	0.011	0.010	0.005
8-13	0.027	0.223	0.478	0.150	0.091	0.068	0.055	0.061
15-20	0.003	0.019	0.081	0.251	0.106	0.055	0.049	0.028
25-32	0.006	0.029	0.138	0.510	0.524	0.461	0.260	0.108
38-43	0.004	0.007	0.023	0.058	0.149	0.293	0.339	0.268
48-53	0.002	0.001	0.004	0.007	0.017	0.055	0.253	0.165
60-	0.001	0.001	0.008	0.007	0.009	0.050	0.134	0.456

6. Conclusion

Automatically classifying the unconstrained age and gender tasks is a challenging research topic while few researchers have paid attention on this issue. In spite of the fact that numerous past techniques have tended to the issues of age and gender grouping, as of not long ago, quite a bit of this work has concentrated on obliged images taken in lab settings. Such settings don't sufficiently reflect appearance varieties normal to this present reality images in social networking sites and online archives. Web images, how-ever, are not just all the more difficult: they are likewise bounteous. The simple accessibility of tremendous image accumulations master videos advanced machine learning based frameworks with viably perpetual preparing information; however this information is not generally suitably named for directed learning. Taking illustration from the related issue of face acknowledgment, experimentation is how well profound CNN+ELM perform on these assignments utilizing Internet information. The results incline with a profound learning architecture designed to keep away from over fitting because of the impediment of constrained marked information. The system is "shallow" contrasted with a portion of the late system designs, along these lines diminishing the quantity of its parameters and the chance for over fitting. The advance swells the extent of the preparation information by falsely including trimmed variants of the images in our preparation set. The subsequent framework was tried on the Adience benchmark of unfiltered images and appeared to fundamentally beat late cutting edge. Two critical conclusions can be produced using the experimental outcomes. In the first place, CNN+ELM can be utilized to give enhanced age and gender arrangement results, not withstanding considering the much little size of contemporary unconstrained image sets named for age and gender classification. Second, the straight forwardness of the model suggests that more involved frameworks utilizing all the more preparing information might well be able to do significantly enhancing results beyond the one reported here.

References

1. T. Ahonen, A. Hadid, and M. Pietikainen. "Face description with local binary patterns: Application to face recognition", *Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041, 2006.
2. S. Baluja and H. A. Rowley. "Boosting sex identification performance", *Int. J. Comput. Vision*, 71(1):111–119, 2007.
3. A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. "Learning distance functions using equivalence relations", In *Int. Conf. Mach. Learning*, volume 3, pages 11–18, 2003.
4. W.L. Chao, J.-Z. Liu, and J.-J. Ding. "Facial age estimation based on label-sensitive learning and age-oriented regression", *Pattern Recognition*, 46(3):628–641, 2013. 1, 2
5. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. "Return of the devil in the details: Developing deep into convolutional nets", *arXiv preprint arXiv:1405.3531*, 2014.
6. J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: "A robust local image descriptor", *Trans. Pattern Anal. Mach. Intell.*, 32(9):1705–1720, 2010.
7. 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 Recognition*, 44(6):1262–1281, 2011. 2
8. T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models", In *European Conf. Comput. Vision*, pages 484–498. Springer, 1998.
9. C. Cortes and V. Vapnik. "Support-vector networks", *Machine learning*, 20(3):273–297, 1995.
10. E. Eidinger, R. Enbar, and T. Hassner. "Age and gender estimation of unfiltered faces", *Trans. on Inform. Forensics and Security*, 9(12), 2014.
11. Y. Fu, G. Guo, and T. S. Huang. "Age synthesis and estimation via faces: A survey", *Trans. Pattern Anal. Mach. Intell.*, 32(11):1955–1976, 2010.
12. Y. Fu and T. S. Huang. "Human age estimation with regression on discriminative aging manifold", *Int. Conf. Multimedia*, 10(4):578–584, 2008.
13. K. Fukunaga. "Introduction to statistical pattern recognition", Academic press, 1991.
14. A. C. Gallagher and T. Chen. "Understanding images of groups of people", In *Proc. Conf. Comput. Vision Pattern Recognition*, pages 256–263. IEEE, 2009.
15. F. Gao and H. Ai. "Face age classification on consumer images with gabor feature and fuzzy LDA method", In *Advances in biometrics*, pages 132–141. Springer, 2009.
16. X. Geng, Z.-H. Zhou, and K. Smith-Miles. "Automatic age estimation based on facial aging patterns", *Trans. Pattern Anal. Mach. Intell.*, 29(12):2234–2240, 2007.
17. B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: "A neural network identifies sex from human faces", In *Neural Inform. Process. Syst.*, pages 572–579, 1990.
18. A. Graves, A.-R. Mohamed, and G. Hinton. "Speech recognition with deep recurrent neural networks", In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE Inter-national Conference on*, pages 6645–6649. IEEE, 2013.

19. G. Guo, Y. Fu, C. R. Dyer, and T. S. “Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression”, *Trans. Image Processing*, 17(7):1178–1188, 2008. 2
20. G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. “A study on automatic age estimation using a large database”, In *Proc. Int. Conf. Comput. Vision*, pages 1986–1991. IEEE, 2009.
21. H. Han, C. Otto, and A. K. Jain. “Age estimation from face images: Human vs. machine performance”, In *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013.
22. T. Hassner. “Viewing real-world faces in 3D”, In *Proc. Int. Conf. Comput. Vision*, pages 3607–3614. IEEE, 2013.
23. T. Hassner, S. Harel, E. Paz, and R. Enbar. “Effective face frontalization in unconstrained images”, *Proc. Conf. Comput. Vision Pattern Recognition*, 2015.
24. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. “Improving neural networks by pre-venting co-adaptation of feature detectors”, *arXiv preprint arXiv:1207.0580*, 2012.
25. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. “Labeled faces in the wild: A database for studying face recognition in unconstrained environments”, Technical re-port, Technical Report 07-49, University of Massachusetts, Amherst, 2007.
26. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Gir-shick, S. Guadarrama, and T. Darrell. “Caffe: Convolutional architecture for fast feature embedding”, *arXiv preprint arXiv:1408.5093*, 2014.
27. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. “Large-scale video classification with convolutional neural networks”, In *Proc. Conf. Comput. Vision Pattern Recognition*, pages 1725–1732. IEEE, 2014.
28. A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Image-net classification with deep convolutional neural networks”, *Neural Inform. Process. Syst.*, pages 1097–1105, 2012.
29. Y. H. Kwon and N. da Vitoria Lobo. “Age classification from facial images”, In *Proc. Conf. Comput. Vision Pattern Recognition*, pages 762–767. IEEE, 1994.
30. A. Lanitis. “The FG-NET aging database, 2002”, Available: www-prima.inrialpes.fr/FGnet/html/benchmarks.html.
31. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. “Back-propagation applied to handwritten zip code recognition”, *Neural computation*, 1(4):541–551, 1989.
32. C. Liu and H. Wechsler. “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition”, *Trans. Image Processing*, 11(4):467–476, 2002.
33. P. Luo, X. Wang, and X. Tang. “Hierarchical face parsing via deep learning”, In *Proc. Conf. Comput. Vision Pattern Recognition*, pages 2480–2487. IEEE, 2012.
34. E. Makinen and R. Raisamo. “Evaluation of gender classification methods with automatically detected and aligned faces”, *Trans. Pattern Anal. Mach. Intell.*, 30(3):541–547, 2008.
35. B. Moghaddam and M.-H. Yang. “Learning gender with support faces”, *Trans. Pattern Anal. Mach. Intell.*, 24(5):707– 711, 2002.
36. X. Niyogi. “Locality preserving projections”, In *Neural In-form. Process. Syst.*, volume 16, page 153. MIT, 2004.
37. A. J. O’toole, T. Vetter, N. F. Troje, H. H. Bulthoff, et al. “Sex classification is better with three-dimensional head structure than with image intensity information”, *Perception*, 26:75–84, 1997.
38. C. Perez, J. Tapia, P. Estevez, and C. Held. “Gender classification from face images using mutual information and feature fusion”, *International Journal of Optomechatronics*, 6(1):92– 119, 2012.
39. P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. “The FERET database and evaluation procedure for face-recognition algorithms”, *Image and vision computing*, 16(5):295–306, 1998.
40. L. Rabiner and B.-H. Juang. “An introduction to Hidden Markov Models”, *ASSP Magazine, IEEE*, 3(1):4–16, 1986.
41. N. Ramanathan and R. Chellappa. “Modeling age progression in young faces”, In *Proc. Conf. Comput. Vision Pattern Recognition*, volume 1, pages 387–394. IEEE, 2006.
42. D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross. “Soft biometrics for surveillance: an overview”, *Machine learning: theory and applications*. Elsevier, pages 327–352, 2013.
43. K. Ricanek and T. Tesafaye. “Morph: A longitudinal image database of normal adult age-progression”, In *Int. Conf. on Automatic Face and Gesture Recognition*, pages 341–345. IEEE, 2006.
44. M. Riesenhuber and T. Poggio. “Hierarchical models of object recognition in cortex”, *Nature neuroscience*, 2(11):1019– 1025, 1999.
45. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. “Image Net Large Scale Visual Recognition Challenge”, 2014.
46. C. Shan. “Learning local binary patterns for gender classification on real-world face images”, *Pattern Recognition Letters*, 33(4):431–437, 2012.
47. Y. Sun, X. Wang, and X. Tang. “Deep convolutional network cascade for facial point detection”, In *Proc. Conf. Comput. Vision Pattern Recognition*, pages 3476–3483. IEEE, 2013.

48. Y. Sun, X. Wang, and X. Tang. "Deep learning faces representation from predicting 10,000 classes", In Proc. Conf. Com-put. Vision Pattern Recognition, pages 1891–1898. IEEE, 2014.
49. M. Toews and T. Arbel. "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion", Trans. Pattern Anal. Mach. Intell., 31(9):1567–1581, 2009.
50. A. Toshev and C. Szegedy. Deeppose: "Human pose estimation via deep neural networks", In Proc. Conf. Comput. Vision Pattern Recognition, pages 1653–1660. IEEE, 2014.
51. I. Ullah, M. Hussain, G. Muhammad, H. Aboalsamh, G. Be-bis, and A. M. Mirza. "Gender recognition from face images with local world descriptor", In Systems, Signals and Image Processing, pages 417–420. IEEE, 2012.
52. V. N. Vapnik and V. Vapnik. "Statistical learning theory", volume 1. Wiley New York, 1998.
53. L. Wolf, T. Hassner, and Y. Taigman. "Descriptor based methods in the wild", In post-ECCV Faces in Real-Life Images Workshop, 2008.
54. S. Yan, M. Liu, and T. S. Huang. "Extracting age information from local spatially flexible patches", In Acoustics, Speech and Signal Processing, pages 737–740. IEEE, 2008.
55. S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang. "Regression from patch-kernel", In Proc. Conf. Com-put. Vision Pattern Recognition. IEEE, 2008.
56. X. Zhuang, X. Zhou, M. Hasegawa-Johnson, and T. Huang. "Face age estimation using patch-based Hidden Markov Model super vectors", In Int. Conf. Pattern Recognition. IEEE, 2008.
57. S.B. Kim , K.S. Han , H.C. Rim , S.H. Myaeng , Some effective techniques for naive Bayes text classification, IEEE Trans. Knowl. Data Eng. 18 (11) (2006) 1457–1466.
58. G.B. Huang , Q.Y. Zhu , C.K. Siew , Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
59. F.S. Khan, J. van de Weijer, R.M. Anwer, M. Felsberg, C. Gatta, Semantic pyramids for gender and action recognition, IEEE Trans. Image Process. 23 (8) (2014) 3633–3645, doi: 10.1109/TIP.2014.2331759 .
60. H. Guang-Bin , C. Lei , S. Chee-Kheong , Universal approximation using incremental constructive feedforward networks with random hidden nodes., IEEE Trans. Neural Netw. 17 (4) (2006) 879–892 .
61. M. Duan, K. Li, X. Liao, K. Li, A parallel multiclassification algorithm for big data using an extreme learning machine, IEEE Trans. Neural Netw. Learn. Syst. PP (99) (2017) 1–15, doi: 10.1109/TNNLS.2017.2654357.
62. B. Zuo , G.B. Huang , D. Wang , W. Han , M.B. Westover , Sparse extreme learning machine for classification, IEEE Trans. Cybern. 44 (10) (2014) 1858–1870.
63. G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybern. Part B Cybern. 42 (2) (2012) 513–529, doi: 10.1109/TSMCB.2011.2168604.
64. Y. Yang , Q.M. Wu , Y. Wang , K.M. Zeeshan , X. Lin , X. Yuan , Data partition learning with multiple extreme learning machines, IEEE Trans. Cybern. 45 (6) (2014) 1463–1475.
65. J. Luo , C.M. Vong , P.K. Wong , Sparse Bayesian extreme learning machine for multi-classification., IEEE Trans. Neural Netw. Learn. Syst. 25 (4) (2014) 836–843.
66. J. Shuiwang , Y. Ming , Y. Kai , 3d convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231.
67. Z. Dong , Y. Wu , M. Pei , Y. Jia , Vehicle type classification using a semi-supervised convolutional neural network, IEEE Trans. Intell. Transp. Syst. 16 (4) (2015) 1–10.