Deep Random Forest for Facial Age Estimation Based on Face Images

Oussama Guehairia, Abdelmalik Ouamane, Fadi Dornaika, Abdelmalik Taleb-Ahmed

Deep Random Forest for Facial Age Estimation Based on Face Images

1st *Oussama Guehairia*
Laboratory of LESIA
University of Biskra
Biskra, Algeria
oussama.guehairia@univ-biskra.dz

2nd *A. Ouamane*
Laboratory of LI3C
University of Biskra
Biskra, Algeria

3rd *F. Dornaika*
University of the Basque Country UPV/EHU
San Sebastian, SPAIN
IKERBASQUE, Basque Foundation for Science
Bilbao, SPAIN

4th *A. Taleb-Ahmed*
IEMN DOAE UMR CNRS 8520 Laboratory
Polytechnic University of Hauts-de-France
Valenciennes France

***Abstract*—The face image of an individual is important for most biometrics systems. The face picture gives loads of helpful informations, including the individual's personal identity, gender, ethnicity, age, emotional expression, and so forth. As of late, a few applications that endeavor age estimation have risen. This paper was aimed to address the problem of image-based human age estimation. It has the following main contributions. We used the advantages of the recent method and algorithms named Gcforest, which proved in several classification tasks, this novel approach includes the power of the decision trees and the advantages of a Cascades structure which allows the interaction between trees. We provide a comparison between two feature types handcrafted and deep feature, we used three databases FG NET, PAL and MORPH II.**

***Index Terms*—age estimation, deep learning, deep feature, transfer learning, random forest,deep random forest.**

I. Introduction

The evolution of the field of biometrics contemporary has become a striking and impressing sight for many and different researchers and scientists. This fact can be viewed as an operative way to update the remarkable progress in technological development in artificial devices, which requires sophisticated and intelligent programs. The biometrics include numerous and diverse fields, the human age estimation considered a recent one. Because of its advantageous job, researchers and specialists are requested to find out about how to appraise ages, as significant assurance prompts, which depict a perceptible job in the nonverbal data to oversite this present reality human – to – human correspondence. It is expressed that the cutting edge movement of the video-based frameworks and social robots has taken research on the keen frameworks as face location, sexual orientation, order, and outward appearance acknowledgment to have the potential outcomes for perceiving and deciphering human ages through time. There has been a developing enthusiasm for the human age estimation process from facial pictures because of an assortment of potential applications. As to assessments, assessing the age of a person from the numerical examination of his face picture is a carefully present topic since it remains a troublesome issue that is impacted by innate and superfluous components. Age estimation by numerical examination of the face picture has diverse potential applications. It can possibly be powerful and helpful for cutting edge video reconnaissance, segment measurement assortment, business knowledge, and client profiling, and inquiry improvement in huge databases. Also, receiving elective approaches to appraise t the human age is found out from different face pictures dependent on the connection between the age data and the facial picture.

The check of the face and advancing the apparatuses utilized in police examinations may be acquired by the utilization of the age quality. Overall, the autoloader age estimation by a machine affects the applications where the point is to foresee the age of a person without identity

In this paper we will present a novel approach in the age estimation task, it will use a new method, which proves it's efficiency in several classification tasks. This method uses several random forests in a cascade structure. The original algorithm of this method named GcForest [1]. GcForest contains two parts Multi Grained scanning which helps in information representation and the cascade information structure which will be named later Deep Random Forest (DRF). Experiments will show the importance of this proposed method. The remainder of the paper is organized as follows: face alignment is briefly introduced in Sect. II. Deep Random Forest. III. Proposed Approach. IV. Experiments and Results. V. we give the conclusion.

II. Deep Random Forest

A. Random Forest

Random forests are strategies for acquiring prescient models for classification and regresion. The technique executes binary decision trees, including CART trees proposed by Breiman et al. (1984). The general idea behind the method is that instead of trying to get an optimized method at once, we generate several predictors before pooling their different predictions. Use the feature to classify or regress a sample of observations described by qualitative and/or quantitative variables. In classification (variable qualitative response): the method predicts the affiliation of observations (observations, individuals) to a class of a qualitative variable, based on quantitative and/or qualitative explanatory variables. In regression (variable continuous response): the method predicts the value taken by a dependent quantitative variable, based on quantitative and/or qualitative explanatory variables. Fig. 1 illustrate the Random Forest classifier, where Each class vector is generated by counting the percentage of different classes of training examples at the leaf node where the concerned instance falls and then averaging across all trees in the same forest.

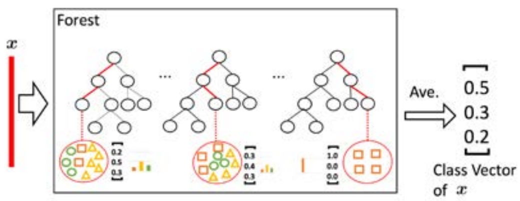

Fig. 1: Illustration of Random Forest classifier.[1].

B. Deep Random Forest

Deep random forest is a novel method proposed by Zhou *et al.* [1], this method considered a good competitor to Convolutional Neural Network in many classification tasks due to the efficacy of decision trees in classification tasks and the way they interact between them. Deep Random Forest shows many advantages, the learning time and the simplicity of the implementation considered the main advantages. DRF is an ensemble of decision trees ensembles tied in a cascade form, inspired by the layer by layer structure in Convolutional Neural Networks Deep Random Forest also composed of many layers where every layer is an ensemble of random forest, each forest of a layer will receive a vector in the input, it will generate a probability class vector which will be concatenated with the original input vector, the dimension of the new vector will be extended as demonstrated in Eq. (1). this newly generated victor will be considered the novel input to the next layer. This structure enables DRF to do representation, learning, The number of cascade levels can be adaptively determined such that the model complexity can be manually set.

$$Dim = D + M \times C \tag{1}$$

where:

- D: the original feature size.
- M: the number of forests.
- C: the number of class.

in the last layer (*Layer L*), the generated probability class vector of each forest will be averaged arithmetically to get one probabilities vectors, in this stage, a max probability index will be chosen as the predicted class. Fig. 2 shows the cascade forest structure. Assume every level of the cascade comprises two (random forest in dark) and two totally decision trees (blue). Assume there are three classes to anticipate, in this way, each forest will yield a three-dimensional class vector, which is then connected for re-representation of the first information.

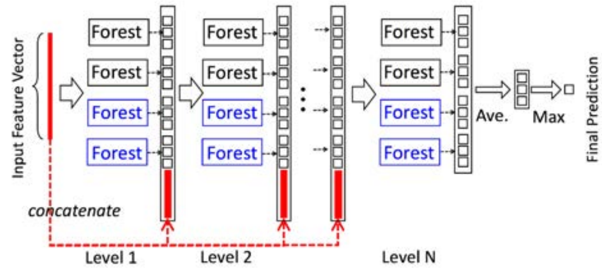

Fig. 2: Illustration of the proposed structure.

III. Experiments and Results

In this work, we stand on the advantages of Deep Random Forest to create our novel method, in which the descriptor vector will be enlarged by concatenating several probabilities class vectors generated by each forest of each layer as shown in Fig. 2. we should elucidate the difference between our work and work in [2]. The main difference is the strategy of training, in our case we used just an ensemble of decision trees but with a special way of interaction between them, where work in [2] switch the training phase between a CNN and differentiable trees.

A. Face alignment

One of the most significant stages in image-based age estimation is Face arrangement. in the presented paper, eyes of each face are distinguished utilizing the Ensemble of Regression Trees (ERT) method [3] which is a powerful and productive calculation for facial landmark localization. When we have the 2D places of the two eyes, we use them to make up for the in-plane pivot of the face. In the wake of playing out the pivot and reusing, the face area ought to be trimmed (adjusted face).

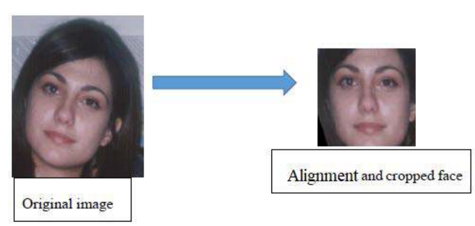

Fig. 3: Face alignment and cropping associated with one original image in PAL database.

B. Face Feature

In this section we will present some of the most used feature for face features extracting, we used three hand-crafted feature and thre deep features to compare the influence of both feature types on the age estimation tasks. as presnt in 8.

1) Local Binary Patterns (LBP): This descriptor was mentioned for the first time in 1993 to measure the local contrast of a image but really popularized three years by Ojala et al. to later analyze the textures. The principle general is to compare the level of brightness of a pixel with the levels of its neighbors[4], [5]. That thus gives an account of information relating to regular reasons in the image, in other words a texture. According to the scale of the vicinity used, certain zones of interest such of the corners or edges can be detected by this descriptor. A texture descriptor can be used by computed LBP labels histograms (each code frequency occurrence) over a region or an image

2) Histogram of Oriented Gradients (HOG): A histogram of Oriented gradient (HOG) is a characteristic used in vision by computer for the target detection. The technique calculates local histograms of the orientation of the gradient on a dense grid, i.e. on zones regularly distributed on the image. It has common points with the SIFT, Shape contexts and the histograms of orientation of contours, but remotely in particular by the use of a dense grid.The local object appearance which shapes the edge directions that can describe an image was the fatal thought that stands behind the historigram of oriented gradients object. The method is particularly effective for the detection of people. The HOG were proposed by Navneet Dalal and Bill Triggs, researchers with the INRIA of Grenoble, conference CVPR of June 2005.

3) Binarized Statistical Image Features (BSIF): This descriptor BSIF was proposed by Kannala and Rahtu (2012) [6], it was used for the recognition of face and the classification of texture. Based on LBP and LPQ, the idea behind the BSIF consists in automatically learning a fixed unit from filters starting from a small whole of natural images, instead of using filters manufactured-with-the-hand like LBP or LPQ. BSIF implies a training, instead of a manual adjustment, to obtain a statistically significant representation of the image, which allows encoder effective information by using the quantification by simple element. The training also provides an easy and flexible manner to adjust the length of the descriptor and to adapt it to the applications presenting of the characteristics of unusual images.

4) Visual Geometry Group (VGG16) Face features: VGG16 is a model of convolutional neural network suggested [7]. The model reaches an accuracy of 92,7% on the basis of data ImageNet [AKR12] which contains more than 14 million images belonging to 1000 classes. It is a model of 16 layers which receives entered size of image RVB of size 224 X 224 VGG16 uses the filters of cores $3X3$ in the layer of convolution (which is smallest cut to capture the concept of right left/, high/low, center) and a filter of size $2X2$ in max pooling, all the hidden layers are equipped with the non-linearity of correction (ReLU). And at the end of the network, three entirely connected layers which calculates the score of each classify characteristics extracted the convolution in the preceding stages. Charts characteristics of the final layer are represented in the form of vectors with scalar values, the function of classification software-max. who formulates the final score between 0 and 1 to present classification penny forms a percentage shared between the various classes, The entirely connected layers are expensive in terms of calculation, of the approaches alternatives were proposed during last years. For example the implementation of a Total layer of pooling average which helps to reduce the number of parameters in network significantly.

5) DEX-IMDB-WIKI and DEX-ChaLearn-ICCV2015 features: The Deep EXpectation (DEX) is a model based in its networks on the VGG-16 architecture, this model pre-trained on ImageNet. What's more, the creators investigated the advantage of fine-tuning over slithered Internet face pictures with available age. In total, they gathered in excess of 500,000 pictures of famous people from IMDb and Wikipedia. The DEX networks were fine-tuned on the crawled pictures and afterward on the gave pictures with apparent age comments from the ChaLearn LAP 2015 challenge on apparent age estimation. We extracted the features by two networks: DEX-IMDBWIKI and DEX-ChaLearn-ICCV2015. The first was prepared on real age prediction utilizing the trimmed and adjusted faces of the IMDB-WIKI dataset, while the subsequent one is a calibrated rendition of the past model, pre-trained on obvious age utilizing the challenge images. A troupe of these models prompted first place at the challenge (115 teams). The uesd features are gathered from the past to the last FC layer.

C. databases

1) MORPH II: This database contains pictures of 13,618 people (guys and females). It contains in excess of 55000 unique of a kind pictures. Every facial picture is commented on with a sequential age. Ages are somewhere in the range of 16 and 77 years. Transform can be separated into more than ethnicity: African, European and other. We utilize the five overlap cross-approval. furthermore, the folds are chosen in such an approach to prevent algorithms from learning the character of the people in the training set by ensuring that all pictures of individual subjects are just in the same fold.

2) FG-NET: This is a widely known database in age estimation. This database has a large variation in lighting conditions, pose and expression. FG-NET contains 1002 facial images associated with 82 individuals. Each individual has more than 10 photos taken in different ages. FG-NET age range is from zero to 69. As in [8], [9], we use the Leave One Person Out cross validation on FG-NET. We leave one individual images out for testing and the other 81 individuals images for training.

3) PAL: The Predictive Aging Lab face is another database from Texas university. It contains 1046 frontal face images (430 males, 616 females). PAL contains faces with different expressions. We perform the random partition as in [10], [11], where we randomly partition images in 80% training and the other 20% for testing. It is repeated five times. The average of the five different splits will be the final performance.

D. Evualuated protocol

In this work we use Mean Absolute Error (MAE), MAE is computed as the average of absolute error between the predicted ages and the ground-truth ones. The MAE is given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P - A|, \quad (2)$$

where N is the number of tested images, P is the predicted age of image i, and A is the ground-truth age of this image.

1) Parameters: This part will take care about experiments, the different results and what we have focused on. At first stage, we should talk about the feature extraction, we used two type of feature extraction hand crafted and deep feature, two model of deep trained already on the age estimation tasks, we used those two models to extract feature vectors from the last fully connected layers as we mentioned above FC6 and FC7. this method known as the transfer learning in deep learning. the use of those two feature vectors aims to have variaus results, which helps later to decide which vector contains rich infomation and should use in this proplem. The input image size to those CNN pretrained model is 224×224. Secondly, we will use the DRF with four forests two random forests and two completely-random tree forests, having 500, 1000 for each kind. the dimension of the deep feature vectors 4096, if we suppose that we have 60 class to predict (noticing that in our case we use every single age as a single class), here every forest will generate a vector of dimension 60. the new dimension will be 4336. DRF layers can be manually chosen. Here we used 1 layer and 2 layer.

Table I: Results obtained with DRF using with one layer.

Descriptor \ Database	FG-NET	PAL	MORPH
LBP	8.81	7.54	8.13
HOG	6.21	5.67	5.20
BSIF	7.89	6.90	8.20
VGG FACE	4.97	4.20	5.20
DEX-CHALEARN	3.90	3.84	4.13
DEX-IMBD	4.02	3.80	4.29

Table I presents the results obtained using the DRF and the FC6 and FC7 feature vectors with one layers,this results obtained by the direct applications of the DRF explained in Fig. 2. this Table shows that FC6 vector is better in term of information then FC7.

Table III: Comparison of our method with the state of the arts.

Method \ Database	FG-NET	MORPH II	PAL
Human workers [12]	4.70	6.30	/
Rank[13]	5.79	/	/
DIF [12]	4.80	/	/
AGES [14]	6.77	8.38	/
IIS-LLD[15]	5.77	/	/
CPNN [15]	4.76	/	/
CA-SVR [16]	4.67	5.88	/
OHRank [17]	4.48	6.07	/
Pontes et al. [18]	4.50	/	/
CAM [19]	4.12	/	/
Rothe et al. [20]	5.01	**3.45**	/
Liu et al. [21]	3.93	/	/
LSDML [8]	3.92	/	/
DRFs [9]	3.85	**2.91**	/
Gunay and Nabiyev [10]	/	/	5.40
Nguyen et al [22]	/	/	6.50
Luu et al [19]	/	/	6.00
Bekhouche et al. [23]	/	/	5.00
Dornaika et al. [11]	/	/	3.79
(DMTL) Hun et al. [24]	/	/	
Structured learning [25]	3.89	/	/
Liu et al.[26]	3.92	3.84	/
Proposed method	**3.82**	4.32	**3.23**

Table II: Results obtained with DRF using two layers.

Descriptor \ Database	FG-NET	PAL	MORPH
LBP	7.61	7.91	9.01
HOG	6.31	5.53	5.63
BSIF	8.02	7.21	8.83
VGG FACE	4.54	4.07	4.57
DEX-CHALEARN	3.89	3.79	4.41
DEX-IMBD	3.82	3.23	4.32

Table II presents the results obtained by the direct use of DRF using two layer, what we can observe that when we passing to the second layer results comes worst then using one layer.

E. Comparison with state of the art

Table III shows the comparison with some of the state of the arts results, our method out perform the state of the arts in FG-NET and PAL but in MOROPH (DRFs) in [9] has the best results.

IV. Conclusion

In this work we used the new method of DRF(Deep Random Forest) that includes the cascade structure using multiple random forest, to rich the input feature vectors with more information. we used both hand-crafted and deep features to compare the influence of both type on our task. results shows that this new proposed method can be a new research topic it concidered a novel technique in this task, training strategy and time complexity prove its efficiency, especially in classification tasks in general and in age estimation specially. it still there many point to focus on it.

References

[1] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," *arXiv preprint arXiv:1702.08835*, 2017.

[2] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. L. Yuille, "Deep regression forests for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2304–2313.

[3] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1867–1874.

[4] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis Machine Intelligence*, no. 12, pp. 2037–2041, 2006.

[5] M. Bereta, P. Karczmarek, W. Pedrycz, and M. Reformat, "Local descriptors in application to the aging problem in face recognition," *Pattern Recognition*, vol. 46, no. 10, pp. 2634–2646, 2013.

[6] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features," in *Proceedings of the 21st international conference on pattern recognition (ICPR2012)*. IEEE, 2012, pp. 1363–1366.

[7] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition." in *bmvc*, vol. 1, no. 3, 2015, p. 6.

[8] H. Liu, J. Lu, J. Feng, and J. Zhou, "Label-sensitive deep metric learning for facial age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 292–305, 2018.

[9] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. Yuille, "Deep regression forests for age estimation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2018, pp. 2304–2313.

[10] A. Günay and V. V. Nabiyev, "Age estimation based on hybrid features of facial images," in *Information Sciences and Systems 2015*. Springer, 2016, pp. 295–304.

[11] F. Dornaika, I. Arganda-Carreras, and C. Belver, "Age estimation in facial images through transfer learning," *Machine Vision and Applications*, pp. 1–11, 2018.

[12] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE Transactions on Pattern Analysis Machine Intelligence*, no. 6, pp. 1148–1161, 2015.

[13] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "A ranking approach for human ages estimation based on face images," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 3396–3399.

[14] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

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

[16] K. Chen, S. Gong, T. Xiang, and C. Change Loy, "Cumulative attribute space for age and crowd density estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2467–2474.

[17] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Computer vision and pattern recognition (cvpr), 2011 ieee conference on*. IEEE, 2011, pp. 585–592.

[18] J. K. Pontes, A. S. Britto, C. Fookes, and A. L. Koerich, "A flexible hierarchical approach for facial age estimation based on multiple features," *Pattern Recognition*, vol. 54, pp. 34 – 51, 2016.

[19] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *2011 International Joint Conference on Biometrics (IJCB)*. IEEE, 2011, pp. 1–8.

[20] R. Rothe, R. Timofte, and L. Van Gool, "Some like it hot-visual guidance for preference prediction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5553–5561.

[21] H. Liu, J. Lu, J. Feng, and J. Zhou, "Group-aware deep feature learning for facial age estimation," *Pattern Recognition*, vol. 66, pp. 82 – 94, 2017.

[22] D. T. Nguyen, S. R. Cho, K. Y. Shin, J. W. Bang, and K. R. Park, "Comparative study of human age estimation with or without preclassification of gender and facial expression," *The Scientific World Journal*, vol. 2014, 2014.

[23] S. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and A. Hadid, "Pyramid multi-level features for facial demographic estimation," *Expert Systems with Applications*, vol. 80, pp. 297 – 310, 2017.

[24] H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen, "Heterogeneous face attribute estimation: A deep multi-task learning approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 11, pp. 2597–2609, 2018.

[25] Z. Lou, F. Alnajar, J. M. Alvarez, N. Hu, and T. Gevers, "Expression-invariant age estimation using structured learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 2, pp. 365–375, 2018.

[26] K.-H. Liu and T.-J. Liu, "A structure-based human facial age estimation framework under a constrained condition," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5187–5200, 2019.