FACIAL AGE ESTIMATION USING TRANSFER LEARNING AND BAYESIAN OPTIMIZATION BASED ON GENDER INFORMATION

Marwa Ahmed[1] and Serestina Viriri[2]

[1]Computer Science, Sudan University of Science and Technology, Khartoum, Sudan
[2]Computer Science, University of KwaZulu-Nata, Durban, South Africa

ABSTRACT

Age estimation of unrestricted imaging circumstances has attracted an augmented recognition as it is appropriate in several real-world applications such as surveillance, face recognition, age synthesis, access control, and electronic customer relationship management. Current deep learning-based methods have displayed encouraging performance in age estimation field. Males and Females have a variable type of appearance aging pattern; this results in age differently. This fact leads to assuming that using gender information may improve the age estimator performance. We have proposed a novel model based on Gender Classification. A Convolutional Neural Network (CNN) is used to get Gender Information, then Bayesian Optimization is applied to this pre-trained CNN when fine-tuned for age estimation task. Bayesian Optimization reduces the classification error on the validation set for the pre-trained model. Extensive experiments are done to assess our proposed model on two data sets: FERET and FG-NET. The experiments' result indicates that using a pre-trained CNN containing Gender Information with Bayesian Optimization outperforms the state of the arts on FERET and FG-NET data sets with a Mean Absolute Error (MAE) of 1.2 and 2.67 respectively.
.

KEYWORDS

Age estimation, Gender information, Deep learning, Convolutional Neural Network, and Bayesian Optimization.

1. INTRODUCTION

The humanoid face carries important information such as expression, gender, age, identity and ethnicity. Age estimation is the determination of the age automatically using a given facial image. It has attracted many researchers' attention, because it has many real-life applications such as surveillance, face recognition, age synthesis, access control, and electronic customer relationship management. The face appearance of a person changes through the growing older process. These changes rise the difficulty of the computer for face recognition tasks. Aging has two stages: (1) Early growth stage which occurs from birth to adulthood, in which there are grander changes in shapes (craniofacial growth), (2) Adult aging stage is from adult to old age as a result of the texture changes (skin aging). These changes in appearance are contributed with several factors like health, race, lifestyle, climate, working environment, increase or decrease in weight, drug use, smoking, emotional stress, and diet [1], [2].

Automatic age estimation for facial images has a series of problems since the age of human varies based on a lot of aspects which may be internal factors such as gender, genetic, etc. [3].

Females and Males have a variable type of appearance aging pattern; this results in age differently [1], [4]. This is caused by the difference in beard and mustaches in males and

hairstyle, makeup, accessories in females. These above facts lead to assume that using gender information may improve the age estimator performance [5]. This paper used Gender Information to advance the age estimation's performance. A CNN is trained for Gender Classification. After training, the resulted CNN is used as an input for age estimation model, which uses Bayesian Optimization to obtain good result through minimizing classification error on the validation set.

This paper uses Gender Information to enhance the age estimation's performance. After getting this information relying on a CNN, Bayesian Optimization is applied to select the best result.

The remaining of this paper is organized as follows. The literature review and related works are introduced in Section 2. Proposed Model is described in Section 3. Experimental results are discussed in Section 4, and Section 5 concludes the paper.

2. LITERATURE REVIEW AND RELATED WORK

2.1. Gender Classification

Automatic gender classification is significant for a lot of applications like targeted advertisements, surveillance etc. It is a task of differentiating between females and males based on the human's features [6]. Females and Males have a variable type of appearance aging pattern; this leads to age differently [1], [4]. This is caused by the difference in beard and mustache in males and makeup, hairstyle and accessories in females.

To solve the gender classification problem [35], numerous techniques used physical appearance as the classification's input. Physical appearance incorporates facial features such as the cheeks, eyes, lips, ears, nose, forehead, hair and the lower and mid body parts, for example, hands, stomach area, legs and so forth. Many study papers have facial features as the classification issue's input. Recently, PC has gotten well known and picking up consideration massively in distinguishing proof of ethnicity of human faces, age and gender, for that reason image processing have a major function in computer learning fields [36]. While diagnosing gender there are some recognizable features that exist among male and female which are utilized by computerized strategies to categorize gender [36].

2.2. Deep Learning

Deep learning is a playing field in machine learning procedures, such that layers of data handling stages are exploited in hierarchical designs for pattern classification and feature learning. It lies in the associations of many study areas, including graphical modeling, neural networks, pattern recognition, optimization and signal processing. The fundamental Deep learnings idea has been created from the investigation of artificial neural system [8]. Convolutional Neural Networks (CNNs) are one of the best significant deep learning approaches whose few layers are trained in a strong manner [10]. The whole picture is convolved in convolutional layers by abusing the moderate feature maps in addition to diverse kernels, creating different feature maps. CNN in general is a progressive neural network, in a way that its pooling layers interchange with convolutional layers [11]. The Pooling layer comes next to the convolutional layer. This layer reduces the feature maps dimensions and network parameters. These layers are interpretation invariant in light of considering calculations of neighboring pixels [10]. The last pooling layer is followed by the Fully-connected layer. For extra feature representation, there are numerous fully connected layers which convert the 2D feature maps to a 1D feature vector. Compared to a traditional neural network, Fully-connected layers achieve similar to the traditional neural network and hold about 90% of the elements in a CNN [6] [10].
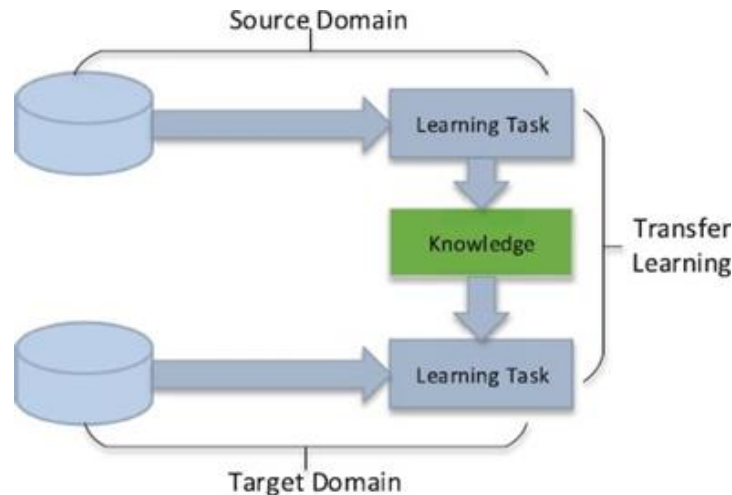
Figure 1: Learning process of transfer learning [7].

Transfer learning: Transfer learning is a significant tool to resolve the basic problem of inadequate training data in machine learning. It transfers the knowledge beginning with a source domain towards a target domain. This will result in a great positive effect on many areas that are challenging to be improved due to insufficient training data [7]. The learning process of transfer learning is displayed in the Fig. 1. In another way; transfer learning can be defined as follows: Assumed that there is a source domain D_S by an equivalent source task T_S besides a target domain D_T by an equivalent task T_T. Then transfer learning is defined as the method of improving the target predictive function f_T (.) relying on the associated information from T_S and D_S, where T_S Ç T_T or D_S Ç D_T [8].

In recent times, many researchers applied deep learning algorithms to face related tasks like gender identification, face verification and age estimation. For face verification process, a Deep ID structure is proposed to extract discriminative features from facial images [12]. Then a verification constraint is added in loss function to improve the Deep ID algorithm to get better performance [13]. [14] Proposed a structure of cascaded Deep ConvNets to discover landmark points of the facial image. Also [15] proposed a new algorithm named deep multi-task learning algorithm to discover landmark points. Centered on the deep learning model, a new framework is built for age feature to be used for the automatic age estimation [16]. The proposed scheme integrates the manifold learning algorithm. Experiments evaluated on two data-sets indicate that their approach is better than the state-of-the-art.

A framework for age estimation based on deep learning is proposed. Due to the lack of labelled images, transfer learning is used. A new loss function is defined for age classification through distance loss addition to cross-entropy loss for relationships description between labels due to the ordered labels in age estimation. Results achieved prove the good algorithm performance in contrast to the state-of-the-art methods [17].

Deep learning is used with a fast and robust age modelling algorithm to propose age estimation system. The local regressor's performance for most groups is indicated as it is better than the global regressor. They evaluate the system on the ChaLearn Looking at People 2016 Apparent Age Estimation challenge data- set, and outcomes in 3.85 MAE on the test set [18].

[19] Introduced the largest public IMDB-WIKI dataset with age and gender labels. VGG-16 architecture for CNNs which are pre-trained on ImageNet data-set is used. A strong face

alignment is done. The apparent age estimation and the perceived age by other humans have been studied. They evaluate the methods on standard benchmarks and results reach state-of-the-art for both apparent and real age estimation. Both the structure innovation and final label encoding are explored [20]. Both global and local signs in aging can be captured by their deep aging feature. Experimental results on the MORPH-II and the FG-NET databases of age prediction show that their proposed deep aging feature outperforms state-of-the-art aging features.

2.3. Bayesian Optimization

Bayesian optimization (BO) is the process of discovering the minimum of a function f(x) on some bounded set X. It constructs a probabilistic model for f(x), which is exploited to provide decisions round where will be the next function evaluation in X, whereas integrating out uncertainty. The major philosophy is to utilize all of the offered information from previous evaluations of f(x) and not only depend on local gradient and Hessian Approximations. This resulted in a technique that can obtain the least possible of difficult non-convex functions through proportion to tiny evaluations, at the cost of functioning more computation to choose which the next point to try is. When evaluations of f(x) are costly to accomplish — as at the state when it requests training a machine learning algorithm as a result — it is easy to warrant some further computation to make better decisions. For extra overview of the Bayesian optimization and a previous work's review, refer to [21].

Algorithm: Bayesian Optimization
1: **for** n= 1, 2… do
2: select new x_{n+1} by optimizing acquition function α
$x_{n+1} = arg_{max}\alpha(x; D^n)$
3: query objective function to obtain y_{n+1}
4: augment data $D_{n+1} = D_n(x_{n+1}, y_{n+1})$ 5: update statistical model
6: **end for**

Mathematically, taking into account the problem of determining a global minimizer (or maximizer) of an unrecognized objective function f

$$X^* = \text{argmax}_{x \in X} \ f(x) \qquad (1)$$

Where X is carefully chosen design area of interest; surrounded by global optimization, X is frequently a compact subset of R_d however the Bayesian optimization's framework can be stratified to additional unusual search spaces that include categorical or conditional inputs, or even combinatorial search areas with many categorical inputs [22].

There are two main selections that must be made when performing Bayesian optimization. A prior over functions must be chosen firstly that prompt assumptions around the function wanted to be optimized. Gaussian process prior, is tractable and flexible. Then an acquisition function must be selected. This function is used to construct a utility function from the model posterior, allowing us to decide which the next point to be evaluated is [23].

2.3.1. Gaussian Processes

The Gaussian Process (GP) is strength-full and suitable on the functions earlier distribution, the functions will be used here to the form f: $X \rightarrow R$. GP is defined by the quality or characteristics which is restricted in size of N points $\{x_n \in X\}_{n=1}^{N}$ Persuade of an assorted Gaussian dispersed

on R_n. The nth of these points is employed to be operated value $f(x_n)$, and the elegant marginalization's properties for the Gaussian distribution authorize us to compute marginals and conditionals in closed form. The support and properties of the distribution that lead to functions are indicated by a mean function and a positive definite covariance function. The support and properties of the resulting distribution on functions are quantified by a positive definite covariance function K: X × X → R and a mean function m : X → R [24].

2.3.2. Acquisition Functions for Bayesian Optimization

Supposing that this function $f(x)$ is drawn from a Gaussian process prior, moreover that, the observations are shaped with $\{x_n, y_n\}_{n=1}^N$, where v and $y_{n \sim N(f(x_{n,v})}$ is the noise's variance proclaimed into the function's observations. These prior and data yield a posterior over functions; the acquisition function, which we denote by $a: X \to R^+$, describes which point in X advised to be assessed next over a proxy optimization $x_{next} = argmax_x a(x)$, where several varied functions have been proposed. In general, these acquisition functions depend on the previous observations as well as the GP hyperparameters; this dependency is signified as $\sigma(x; \{x_n, y_n\}, \theta)$. Acquisition functions have a numerous popular options. Under the Gaussian process prior, these functions depend on the model exclusively above predictive variance function $\mu(x; \{x_n, y_n\}, \theta)$ and its predictive mean function $\sigma^2(x; \{x_n, y_n\}, \theta))$. In the proceeding, the best current value is signified as $x_{best} = argmin_{xn} f(x_n)$ in addition to the cumulative distribution function of the standard normal as $\emptyset(.)$ [24].

2.3.3. Probability of Enhancement

One instinctive strategy is to increase the improving possibility above the best current value [13]. This can be calculated systematically under the GP as:

$$\sigma_{P1}(x, \{x_n, y_n\}, \theta) = \emptyset(y(x)), \quad y(x) = \frac{f(x_{best}) - \mu(x, \{x_n, y_n\}, \theta)}{(\sigma, \{x_n, y_n\}, \theta)} \tag{2}$$

2.3.4. Expected Improvement

Else, one could decide on to increase the expected improvement (EI) over the current best. This correspondingly has closed form under the Gaussian process [24]:

$$\sigma_{E1}(x, \{x_n, y_n\}, \theta) = \sigma\left(x, \{x_n, y_n\}, \theta)(\gamma(x)\emptyset(\gamma(x)) + N(\gamma(x); 0,1)\right) \tag{3}$$

2.3.5. GP Upper Condence Bound

An extra recent development is to exploit lower confidence bounds (upper, when dealing with maximization) to build acquisition functions that reduction regret over the course of their optimization [34]. These acquisition functions have the form:

$$\sigma_{LBC}(x, \{x_n, y_n\}, \theta) = \mu(x, \{x_n, y_n\}, \theta) - k\sigma(x, \{x_n, y_n\}, \theta) \tag{4}$$
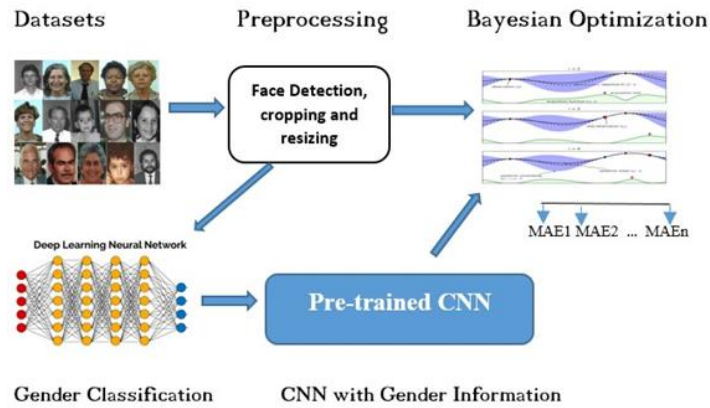
Figure 2: Framework for the Proposed System.

3. PROPOSED MODEL

With the purpose of training CNN, the architecture of CNN has been identified firstly. This CNN is trained for Gender Classification on the datasets. The options used for training this CNN must be stated initially for age estimation procedure. The technique of selecting and fine-tuning these training parameters is challenging and time-consuming. Bayesian Optimization (BO) algorithm is suitable to optimize internal parameters of classification and regression models. It can be used to optimize functions that are discontinuous, time-consuming and non-differentiable to assess. A GP model of the objective function, which is used to assess this model, is maintained internally. BO is applied to the pre-trained network to find optimal training options. . Fig. 2 shows the framework of our proposed system, beginning by pre-processing images, which includes Face detection, Face cropping and face resizing. The pre-processed data is used by the CNN for gender classification purpose. BO uses the pre-processed data and the resulted pre-trained network with Gender Information. This pre-trained network is fine-tuned for age estimation task. Training and testing this model give a number of results (n) which is equal to the objectives' number that stated in the BO. As a final point, the best result that has been achieved from these n results is chosen as the final result to be accepted. The details of our proposed system is clarified by the next steps.

1. Pre-processing stage for all facial images.
2. Training CNN to get Gender Information.
3. Identify the variables for optimization using Bayesian Optimization.
4. State the objective function.
5. Perform Bayesian optimization.
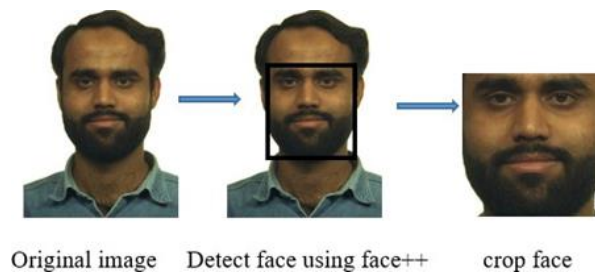6. Select the best network to be loaded.

Figure 3: Sample Image before and after Pre-processing.

3.1. Dataset Pre-Processing

Firstly, Face++ Detector [25] is applied for all the images to detect faces. The faces has been extracted without any noise like; hair and other features. **Fig. 3** displays the preprocess stage on a sample image for the two datasets. This phase resulted in having two types of images for each subject: the detected image and the original one. As a final point, a resize process is done for these detected face to 80x80 to be suitable with the training on the network.

3.2. Training CNN to get Gender Information

Our Network consists of seven convolution layers. The options of training have been specified firstly. The pre-processed facial images have been used for training to classify Males and Females. Finally, the resulted network has been saved to disk after the training completed. It has been saved to be fine-tuned for age estimation task.

3.3. Select Variables for Optimization

The selected variables for optimization are: Momentum and Initial Learning Rate as shown in Table 1. The best initial learning rate relies on the network used and the data. These variables are the options of the training algorithm.

Table 1: The selected variables for optimization.

Momentum	Initial Learning Rate
[0.8 - 0.95]	[1e-3 - 1e-1]

3.4. Objective Function for Optimization

The objective function is defined. It takes the pre-trained network, values of the optimization variables, and the data as inputs. This function states the training options. Also it involves training and validating this pre-trained network. For each time that the training is completed, the trained network is saved to a disk.

3.5. Bayesian Optimization

Bayesian optimization reduces the classification error on the validation set. The objective function trains the pre-trained network and yields the classification error on the validation set. The best model is selected by using the error rate. To estimate the Mean Absolute Error (MAE), the last selected model is tested on the test set. The maximum number of objective function evaluations is determined by 40 and eight hours are chosen for the two datasets as the maximum total optimization time. All the training and testing is done on a single GPU (NIVIDIA GeForce 840M) with 2GB RAM.

3.6. Final Network Evaluation

The best network has been loaded and evaluated on the test set to get the MAE result for evaluating the final network purpose.

4. RESULTS AND DISCUSSIONS

Performance of Bayesian optimization with deep learning based on Gender information was evaluated by testing its ability to estimate ages.

4.1. Datasets

This experiment uses two facial benchmark datasets as shown in Table 2 for facial age estimation. FG-NET [26] is the first one. It consists of 1002 images with ages amongst 0-69. However, more than 50% of the subjects' ages are between 0 and 13. FERET[27] is the second dataset which is used by many researchers for age estimation. It contains 2250 facial images of 994 subjects. The age range is between 10 and 70.

Table 2: The Datasets used for evaluation.

Name Age	Total Images	Total Subjects	Range
FG-NET	1,002	82	[0 - 69]
FERET	2366	994	[10 - 70]

4.2. Evaluation Metrics

Mean Absolute Error (MAE): We utilizes MAE to assess the presentation of the age estimation model. MAE is utilized to quantity the error between the predicted age and ground truth. It is calculated in equation (5), wherever y^j and y refers to ground-truth and predicted age value, correspondingly, and N demonstrates the quantity of the testing samples.

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} |y_i' - y_i| \qquad (5)$$

4.3. Results Comparisons

As in previous works, Leave One Person Out (LOPO) evaluation protocol is employed for FG-NET dataset. Facial images of one person have been selected randomly for validation, while the remaining images for other subjects have been used for training. This technique is repeated for 82 folds to evaluate the proposed model. The average of the 82 folds' results is eventually accepted as the final result for the proposed model. The quantitative results for FG-NET are summarized in Table 3. As displayed in the table, our model has achieved superior results compared to all the current state-of-the-arts. As mentioned in DLBO [24], the Deep learning is applied with Bayesian Optimization and this results in improving the performance compared to the previous works. In this paper we extend [24] through using DLBO with Gender Information to test if using Gender Classification can improve age estimation. The result in the table indicates that using Gender Information has improved the age estimation model by increasing the performance from 2.88 to 2.67.

Table 3: Comparison of Maes With A State-of-the-Art Approaches
on the Fg-Net Dataset.

Deep Learning-Based Methods	MAE
DEX [18]	4.63
Ranking-CNN [27]	4.13
GA-DFL [28]	3.93
DRFs [29]	3.85
ODL + OHRanker [29]	3.89
ODL [29]	3.71
DLBO [24]	2.88
Proposed	**2.67**

A 10-Fold cross-validation is performed for FERET dataset as in [30]. Specifically, the entire dataset is divided into ten equally sized folds. Nine of these folds utilized for training and the former fold for testing. This process is recurring ten times and the final age estimation result is taken as the average of these ten results. The quantitative results are summarized in **Table 4**. As shown in the table, our proposed model has achieved superior results compared to all the prior state-of-the-art results on FERET. Compared to DLBO [24], using Gender Information increases the performance from 1.3 to 1.2.

Table 4: Comparison of Maes with a State-of-the-Art Approaches
on the Feret Dataset.

Hand-crafted Methods	MAE
MAP [30]	4.87
HAP [31]	3.02
MAR [32]	3.0
Deep Learning-Based Methods	**MAE**
DLBO [24]	1.3
Proposed	**1.2**

5. CONCLUSION AND FUTURE WORK

In this paper Gender Classification is applied on deep learning to get benefits from the Gender Information on age estimation field. Then Bayesian Optimization is used for the resulted pre-trained network to select the optimal training options for Age estimation. The experiments' results show that using gender information with Deep learning and Bayesian Optimization achieves\ good results associated with the state of the arts: outperforms the state of the arts on FERET and FG-NET datasets with a MAE of 1.2 and 2.67 respectively. Also compared to DLBO [24], the experiments illustrate that using gender information improve the age estimation task. Future works involve evaluating this model on MORH dataset which needs better GPU, using images with larger size, images alignment and data augmentation.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

[1] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 11, pp. 1955–1976, 2010.

[2] E. Patterson, K. Ricanek, M. Albert, and E. Boone, "Automatic representation of adult aging in facial images," in Proc. IASTED Intl Conf. Visualization, Imaging, and Image Pro- cessing, pp. 171–176, 2006.

[3] M. Ahmed and S. Viriri, "Age estimation using facial images: A survey of the state-of-the-art," in 2017 Sudan Conference on Computer Science and Information Technology (SCCSIT), pp. 1–8, IEEE, 2017.

[4] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally adjusted robust regression for human age estimation," in 2008 IEEE Workshop on Applications of Computer Vision, pp. 1–6, IEEE, 2008.

[5] N. Lakshmiprabha, J. Bhattacharya, and S. Majumder, "Age estimation using gender information," in International Conference on Information Processing, pp. 211– 216, Springer, 2011.

[6] G. Trivedi and N. N. Pise, "Gender classification and age estimation using neural networks: A survey," International Journal of Computer Applications, vol. 975, p. 8887.

[7] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in International conference on artificial neural networks, pp. 270–279, Springer, 2018.

[8] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big data, vol. 3, no. 1, p. 9, 2016.

[9] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in Proceedings of the 22nd ACM international conference on Multimedia, pp. 157–166, ACM, 2014.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradientbased learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] M. D. Zeiler, Hierarchical convolutional deep learning in computer vision. PhD thesis, New York University, 2013.

[12] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1891–1898, 2014.

[13] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in Advances in neural information processing systems, pp. 1988–1996, 2014.

[14] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3476–3483, 2013.

[15] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in European Conference on Computer Vision, pp. 94–108, Springer, 2014.

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

[17] Y. Dong, Y. Liu, and S. Lian, "Automatic age estimation based on deep learning algorithm," Neurocomputing, vol. 187, pp. 4–10, 2016.

[18] F. Gurpinar, H. Kaya, H. Dibeklioglu, and A. Salah, "Kernel elm and cnn based facial age estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 80–86, 2016.

[19] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," International Journal of Computer Vision, vol. 126, no. 2-4, pp. 144–157, 2018.

[20] J. Qiu et al., "Convolutional neural network based age esti- mation from facial image and depth prediction from single image," 2016.

[21] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with appli- cation to active user modeling and hierarchical reinforcement learning," arXiv preprint arXiv:1012.2599, 2010.

[22] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," Proceedings of the IEEE, vol. 104, no. 1, pp. 148–175, 2016.

[23] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in Advances in neural information processing systems, pp. 2951–2959, 2012.

[24] M. Ahmed and S. Viriri, "Deep learning using bayesian opti- mization for facial age estimation," in International Conference on Image Analysis and Recognition, pp. 243–254, Springer, 2019.

[25] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 386–391, 2013.

[26] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the fg-net ageing database," Iet Biometrics, vol. 5, no. 2, pp. 37–46, 2016.

[27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," IEEE Transactions on pattern analysis and machine intelli- gence, vol. 22, no. 10, pp. 1090–1104, 2000.

[28] S. Chen, C. Zhang, and M. Dong, "Deep age estimation: From classification to ranking," IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2209–2222, 2018.

[29] 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.

[30] H. Liu, J. Lu, J. Feng, and J. Zhou, "Ordinal deep learning for facial age estimation," IEEE Transactions on Circuits and Systems for Video Technology, 2017.

[31] C.-C. Ng, M. H. Yap, N. Costen, and B. Li, "Will wrinkle estimate the face age?," in Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, pp. 2418–2423, IEEE, 2015.

[32] C.-C. Ng, M. H. Yap, Y.-T. Cheng, and G.-S. Hsu, "Hybrid ageing patterns for face age estimation," Image and Vision Computing, vol. 69, pp. 92–102, 2018.

[33] C.-C. Ng, Y.-T. Cheng, G.-S. Hsu, and M. H. Yap, "Multi- layer age regression for face age estimation," in Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on, pp. 294–297, IEEE, 2017.

[34] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," arXiv preprint arXiv:0912.3995, 2009.

[35] Narvekar, Sneha, and J. A. Laxminarayana. "Approaches to Gender Classification: A Survey."2020,.

[36] Salma M. Osman Mohammed, Serestina Viriri, "Gender Identification from Facial Images: Survey of the State-of-the-art", Sudan Conference on Computer Science and Information Technology (SCCSIT), 2017.