

Transfer and Ensemble Learning in Real-time Accurate Age and Age-group Estimation

Anh-Thu Mai^{*1,2}, Duc-Huy Nguyen^{1,2}, Thanh-Tin Dang^{1,3}

¹Vietnam National University Ho Chi Minh City (VNU-HCM), Vietnam

²Training Program of Excellent Engineers in Vietnam (PFIEV), Ho Chi Minh city University of Technology (HCMUT), Vietnam

³Faculty of Electrical and Electronics Engineering, Ho Chi Minh city University of Technology (HCMUT), Vietnam

ARTICLE INFO

Article history:

Received: 22 August, 2022

Accepted: 04 December, 2022

Online: 20 December, 2022

Keywords:

Age Estimation

Convolutional Neural Networks

Transfer Learning

Ensemble Learning

ABSTRACT

Aging is considered to be a complex process in almost every species' life, which can be studied at a variety of levels of abstraction as well as in different organs. Not surprisingly, biometric characteristics from facial images play a significant role in predicting human's age. Specifically, automatic age estimation in real-time situation has begun to affirm its position as an essential process in a vast variety of applications. In this paper, two approaches are addressed as solutions for such application: prediction of accurate age and age group by using the two most fundamental techniques in the domain of deep learning – convolutional neural networks (CNNs) and deep neural networks (DNNs). In summary, this work can be split into two main key contributions. By applying a novel hierarchical aggregation built on the base of neural network developed from the training dataset, in the first stage, features extraction, the convolutional activation features are extracted from the captured facial image. As soon as this part is done, the features classification step is performed, in which Softmax Regression (SR) and majority vote classifiers are applied to predict accurate age and age group respectively. The effectiveness of the designed model was showed satisfactorily in the experimental results, which emphasizes the promising of the solution and indicates another direction for future development of algorithms and models in the field of machine learning.

1. Introduction

Automatic age prediction from face images can be employed in a huge number of scenarios, from video surveillance [1], health-care system, health management and statistics purposes [2], performance of individuals recognition [3], advertisement according to users' age, or even criminal investigations [4]. The problem is the reliability of the results since the acquired images could be in unconstrained conditions because of lighting or pose, or when the photographed individuals do not aware of the cameras' presence. Thanks to the advent of computational approaches, these constraints become less and less stressful for automatic computer vision algorithms. Furthermore, this analysis must be completed in a limited time interval of milliseconds, owing to the fact that it is performed in real time condition. Consequently, it is necessary for the solution of this task to be both fast and accurate.

In order to achieve this goal, first of all, face detection is performed using a tool in real-time computer vision's deep learning based-face detector, OpenCV. Particularly, this method is built mainly on ground of Single Shot Detector (SSD) and ResNet backbone for the purpose of increasing the accuracy and accelerating the estimation process, which could be considered to be a compensation for some acceptable major setbacks such as having unconscious biases in the training set or detecting darker-skinned people less accurately than lighter-skinned ones when comparing with Haar-likes cascades [5] or histogram of oriented gradients (HOG) [6]. After the previous step has been fulfilled, by training with an abundant dataset, an age prediction model is created, thanks to which the features corresponding to each age are extracted using a structure of convolutional neural networks (CNNs). Meanwhile, a binary large object (blob) constructor pre-processes the input images and extracts the face region of interest (ROI) so as to calibrate the convolutional architecture for fast feature embedding (caffe) based model that is utilized in this work.

*Corresponding Author: Anh-Thu Mai, Email: thu.mai.k18@hcmut.edu.vn

Table 1: Outline of Some Outstanding Studies on Age Prediction

Publication	Feature representation	Face database/Database size	Algorithm	Evaluation protocol	Performance/Accuracy
[14]	2D shape, raw pixel values	Private/500	Regression	500 train, 65 test	MAE 4.3
[15]	Anthropometric model	HOIP	Classification	Leave-one-out	CS 57.3%(M), 54.7%(F)
[16]	Ages pattern subspace (AGES)	FG-NET/1002 MORPH	Regression	Train on FG-NET, Test on MORPH	MAE 8.83; CS 70%
[9]	Active appearance model	FG-NET/1002	Hybrid	Leave-one-person-out	MAE 4.97; CS 88%
[17]	BIF	MORPH II/55,132	Regression	50% train, 50% test	MAE 4.2
[11]	Active appearance model, label distribution	FG-NET/1002 MORPH II/55132	Classification	Leave-one-person-out (FG-NET), 10-fold cross-validation	MAE 4.8(FG-NET), 4.8(MORPH II)
[18]	Kullback-Leibler/raw intensities	LFW/150,000 MORPH II/55,132 FG-NET/1002	Classification (CNN)	N/A	MAE 2.8(FG-NET), 2.78(MORPH)

The remaining of this paper is arranged as follow. In the next section, we will briefly review some researches inspiring our work. The architecture of the model will be explained precisely in section 3. The overviewed algorithms and the detailed experimental results will be discussed in section 4 and 5 respectively. The last section provides a summary of this work as well as offers a conclusion and possible future development.

2. Related works

Since the early 2000s, automatic age prediction has drawn researcher's attention in the field of machine learning and has undergone a renewed interest some years later as a consequence of the availability of larger databases having more real-annotated date than in the past. Some noticeable researches are shown in Table 1 to represent the evolution of the methods. Overall, the process of this task involves two main phases: features extraction to pick out particular aging characteristics and features classification to divide these features.

Feature extraction stage. The use of a classifier based on support vector machines (SVMs) and support vector regression (SVR) is widely used in several distinct works, whereas other researchers, who concentrated on achieving a faster and better performance, relied mainly on the combination of textural and local appearance. Facial images analysed by those methods will be capture the variation in shape and intensity. Nevertheless, the applied of deep learning algorithm, especially CNNs and Deep Neural Networks (DNNs) in age estimation has become more and more popular [7][8]. Particularly, the work of using ordinal regression and multiple output CNNs for this task by Niu. has demonstrated superior performance in comparison with other methods, and acts as motivation factor for the development our the deep learning approach.

Pattern classification. It has been proposed, including, among others, to use SVMs and SVR [9], Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA) [10], along with their regularized and kernelized versions, neural networks and their variant of Conditional Probability Neural Network [11]. In multi-task classification issues, Softmax Regression (SR) [12] shows that

it outperforms other methods in terms of performance. A detailed examination on these aging classification approaches has been shown in [13].

3. Deep learning based-face detector

According to the aforementioned technique that is used in this paper, the first step that is performed is face detector using OpenCV. More precisely, the performed face detector is established on a SSD following a ResNet-like features which allow the result to be both faster and more accurate. Besides, this method uses DNNs module barked into OpenCV's library without adjusting parameters and build a lightweight output model. Since the face detection task is much more accurate for lighter-skinned people, we make a decision of calibrating the pre-trained face detector with a dataset of multi-ethnic so as to regain the balance among different skin's colors.

4. Accurate age prediction

Before building the model from pre-trained weights, compiling it with some functions and then training, the facial images from the dataset were pre-processed. Figure 1 shows the overall approach.

4.1. IMDB-WIKI dataset

Among online available database, e.g. MORPH, FG-NET, Kaggle, etc. we chose public celebrities' images from IMDB-WIKI dataset, which offers approximate 0.5 million of images ranging from 0 to 100 years old. The age distribution of this dataset is shown in Figure 2. Furthermore, in this process, several images were randomly ejected so as to each age has the same amount of images. Therefore, about 62,000 images remain for training our CNNs.

4.2. Model's construction using VGG-like pre-trained weights:

In purpose of strengthen the quality of distinguished features, prior to training, a histogram equalization is applied for facial images. This pre-processing stage has claimed its importance from the received results.

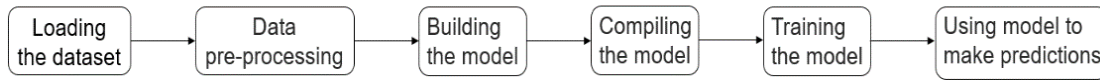


Figure 1: Ensemble process

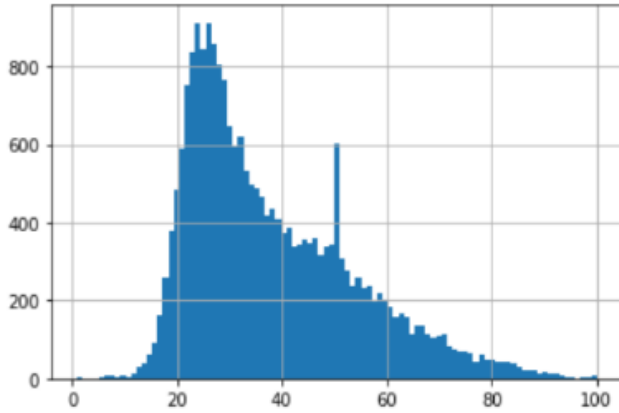


Figure 2: Age distribution histogram from the IMDB-WIKI dataset

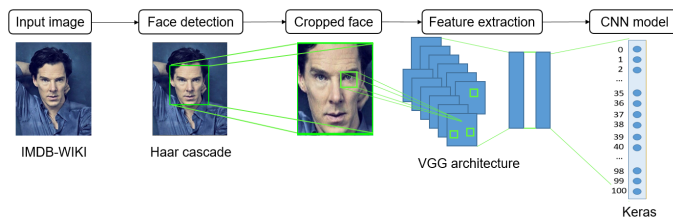


Figure 3: A general description of our procedure for developing CNN architecture in Keras

In this paper, we use VGG-like pre-trained weights and build our VGG model to deal with the experiments, which grants us the opportunity to fine-tune only the weights of the final layer instead of building and training from the scratch. The essential stage in the compilation process is believed to be selecting the optimizer and loss function. We employ "Adam", i.e. adaptive moment estimation, to adjust the learning rate in the performance of the optimization step. A lower learning rate would result in more accurate weights but in the interim it will stretch the handling time for other weights. Whereas for our loss function and accuracy calculating apparatus for the data use for validating, we utilize Categorical Cross Entropy [19], the foremost common choice for classifying patterns. As specified, transfer learning scheme was applied to train the model to assure that we obtain an output of the overall accurate age combined from 101 layers. Consequently, we might have an outcome with a superior achievement on human face's biometric characteristics.

When processing, an image with a fixed size 224x224 RGB was inputted into this network. The model has five convolutional blocks with a max-pooling at the end of each block. The number of filters are orderly 64, 128, 256, 512 and 512. A block with three dense layers is inserted at the end of the model. The final dense layer acts as a softmax function. This model type used is Sequential which is the most straightforward way to construct such model

layer by layer in Keras, which totally matches with the VGG neural network architecture [20]. The process of building such model's substructure architecture is clearly shown step-by-step in Figure 3.

In order to address a multi-class classification task, as aforementioned, the classifiers that show the finest results are usually a CNN which has SR as the last layer. The previous layers are a combination of Convolutional layers and nonlinear activation functions and pooling. These layers can be considered as a tool to help extract features of the data, the last layer is SR, a transparent but exceptionally effective linear classifier. By this way, we are able to assume that a collection of one-vs-rest classifiers is trained simultaneously, supporting each other, so it will result in better performance than training separately each classifier. The adequacy of CNNs, is that the both models use for feature extraction and classification will be trained together. This classifier allows us to find feasible coefficients which pair with the found feature vector, while the role of the feature extraction model is to fine-tune the coefficients of the convolutional layers in order to receive a linear, consistent feature result in comparison with the classifier in the last layer. We also make a checkpoint to monitor model over iterations and prevent it from overfitting. The iteration which has the least validation loss value will incorporate with the optimum weight coefficients so the validation loss will be monitored and only the best result will be saved.

5. Age-group prediction

5.1. Adience dataset

For age group estimation, Adience benchmark dataset, one of the newest databases which is designed for this kind of task, is used. Over 16000 unconstrained facial images are divided into 8 different age groups: 0-2; 4-6; 8-13; 15-20; 25-32; 38-43; 48-53; 60-.

5.2. Model using Caffe-based

Fistly, the input data images will have to go through a pre-processed phase, in which they will be randomly resized and cropped, subtracted mean values, scaled values with a scaling factor, switched the channel Blue to Red and vice versa by a function from OpenCV. As a result, we achieve a 4-dimensional main matrix with data formatted in batch size, channel, height, and width. The whole detailed process can be seen below in Figure 4.

The pre-trained caffe-based model, which used for detecting ages will go through a calibrating process. In this designed network, the system is built with three convolutional layers, each taken after by a rectified linear unit (ReLU) and followed by a pooling layer. The primary two layers are applied normalization by the local response normalization [21]. Constructed right after these convolutional blocks, a collection of 512 neurons will create a fully-connected layer and there are two such kind of layers.

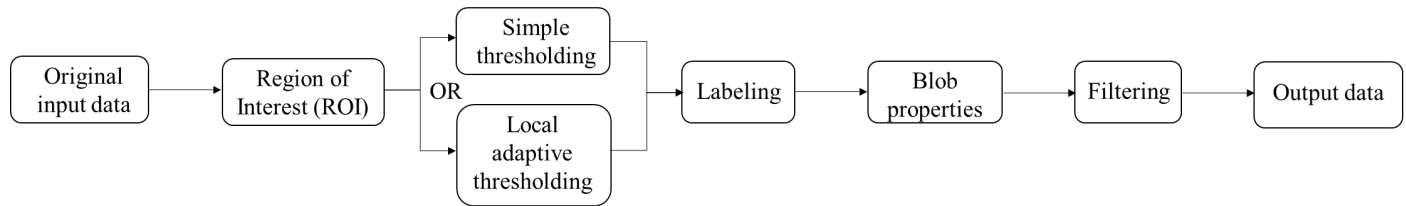


Figure 4: Data pre-processing pipeline

We tested with the method of using the designed model so as to predict age-group for novel faces. In fact, each 256×256 human’s facial image will produce an assemblage of four 227×227 which are cropped from each edge as well as another one which takes the center position of the original image as the center. The network is represented by all five of these images in combination with their horizontal reflections, and the final estimation is chosen to be the average value over all of the variants. Nonetheless, there are numerous challenges with the Adience dataset’s images, like occlusions or motion blur, which can cause minor misalignments and have a noticeable effect on the accuracy of the calculation. The over-sampling method is designed for this situation. By bypassing the requirement to improve alignment quality and focusing on immediately supplying the network with many translated versions of the same face, it can correct these misalignments.

5.3. Transfer Learning on VGG Architecture for age-group prediction

In Figure 5, we illustrate our pipeline for the proposed architecture that is designed and realized in this work. The convolutional layers are common and follow the pattern of those of VGG16 architecture. The fully connected layers, however are reduced because of our eight target classes. Hence, fc6, fc7 and fc8 has 512, 512 and 8 neurons, respectively.

Transfer learning: The weights of the convolutional network are loaded from VGG model trained on VGG-Face dataset. The model is then fine-tuned. Because there are so few training data, this is a vital procedure. This also provides quicker convergence and reduces overfitting for the model. The network is trained using data from the OUI-Adience Face Image Project. The main steps are outlined in Figure 5, where a pre-trained VGG-Face model based on the two datasets: IMDB-WIKI and VGG-Face CNN descriptor is loaded, froze superfluous weights, and added a customized classifier with pointed layers of trainable parameters to improve the original configuration. This customized classifier is then added to solve the multi-class classification task by training classifier layers on targeting training data, adjusting hyper-parameters, and unfreezing some additional layers if needed.

5.4. Ensemble learning

The next phase involves putting both of the aforementioned models into practice simultaneously under fresh conditions and selecting the superior model by evaluating the outcomes. The category that receives the most votes becomes the main age when classifying an image, which gives the impression that it is an election. As a result, we decide to combine the results using both the weighted majority voting and the soft voting. A summary of the process is shown in Figure 6.

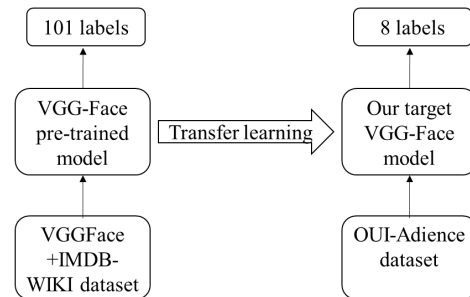


Figure 5: Overview of the Keras-based pipeline for CNN models

6. Overviewed algorithm

6.1. Accurate age prediction algorithm from front view

Because the change of face appearance is less changing and contains more information in frontal view than in profile views with the same rotation in depth, we only construct an age predictor for frontal view with a slight rotation.

When making a prediction from a new face image, the outputs will be a column tensor in which each element corresponds to a class’s score. These scores, through softmax, will be normalized into probabilities and softmax classifier has a probabilistic interpretation [22] which will assign the probabilities to each class.

An input x has a probability a_i that it would fall into class i -th. For the age-set problem, y_i is a label, f_j represents the j -th element of the vector of class scores f , which connects the raw image through some rules to the class scores. By combining them, we get the following functions: the softmax function (1) and the cross-entropy loss function (2) [23]:

$$a_i = \frac{e^{f_i}}{\sum_j e^{f_j}} \quad (1)$$

$$L_i = -\log\left(\frac{e^{f_i}}{\sum_j e^{f_j}}\right) = -\log(a_i) \quad (2)$$

The result of this calculated prediction would be inaccurate and unreliable if we choose the one with highest possibility as the output, as there are a plenty of different classes in this task to estimate accurate age. In order to make this prediction become more reliable, in other words convert this task into a regression classification task, we calculate the probability of each class having an accurate age. By multiplying each output by its corresponding label and computing the sum of those values, we arrive at a summed result. Using the softmax expected value, E , as

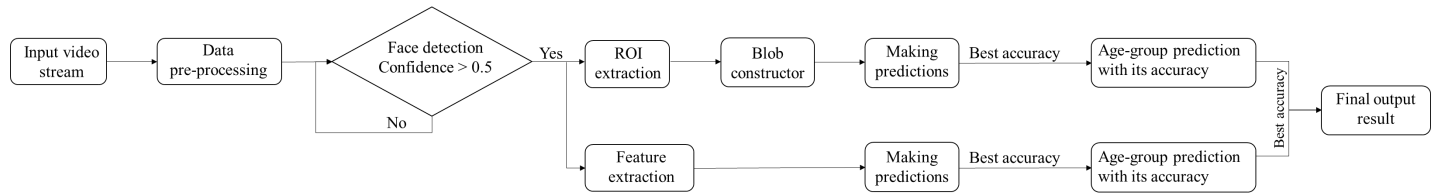


Figure 6: Our overall method

the basis for our apparent age prediction, we calculate the apparent age prediction as follows:

$$E(A) = \sum_{i=0}^{100} y_i a_i \quad (3)$$

6.2. Majority vote classifiers

Age-group predictions rely primarily on majority votes. In terms of results fusion, among all hard voting and soft voting are the most common schemes.

Hard voting, also known as weighted majority voting, is defined based on the assumption that every image x has the classification $C(x)$ of a model i -th, whose weights w_i , with the total of w_1, \dots, w_m is equal to 1, C is from an collection of n labels along with $I(\cdot)$ as an indicator function. Each classifier will vote for a class, and the majority class \hat{y} with most votes [24]:

$$\hat{y} = \arg \max_{j \in n} \sum_i^m w_i I(C_i(x) = j) \quad (4)$$

Soft voting, on the other hand, equalizes all classifiers to provide its proper probability for a particular target class. Supposing an image x receive the predicted result \hat{y} , with m is the number of models and n is the number of classes, then the probability is p_{mn} . The label which possesses the highest sum of weighted probability will represent as the output for this predicting method:

$$\hat{y} = \arg \max_{j \in n} \frac{1}{m} \sum_i^m p_{ji}(x) \quad (5)$$

6.3. Real-time application

Face detection is performed by using the Feature Base Approach method. It locates faces by extracting structural features from some facial regions like eyes, nose, mouth etc., which help distinguish between a face and many other objects and then uses them to detect a face. In our paper, we implement a feature-based approach by using OpenCV.

Instead of creating and training the model from scratch, we use one of those trained Haar Cascade models in OpenCV which are available as XML files. In developing this application, we applied our model for real-time situation. Due to the fact that a real-time video stream is an assemblage of images, face detection is then applied for each frame taken by computer. By combining this model with age prediction, we then carry out the estimation step and complete the whole model.

7. Experimental results

In Figure 7 and Figure 8 below, we have illustrated two real-time test cases where each consists of three cases which are Age-group prediction only (left), Accurate age prediction only (center) and Simultaneous age prediction (right): single face and standard condition, and the other, multiple faces and unfavorable condition. Based on this typical case, we discover that our final model gives the most accurate and stable results with the actual age of the sample compared to the two monotone models that we originally built individually. In fact, the efficiency of the age prediction model is its association algorithm, because no matter how good the dataset is, it cannot wipe out all input situations (within the collection ability of our group). Therefore, we choose the optimal development of the algorithm rather than building the perfect dataset from scratch.

7.1. Accurate age prediction

The selective dataset was split into around 16000 train instances and then 6000 or thereabouts test instances. Our final train loss and accuracy were 2.892846 and 0.239473, respectively. Based on 6000 images of the test instances, our target model's mean absolute error (MAE) reaches the value of 4.09 years.

There are some typical scenarios in Table 2 in which we test our modes in real-time conditions and then evaluate their performance. We showed the summary of 50 example cases and each accurate age was tested by 5 different people. The average accuracy is 91.97%, age difference is 2.7 years, with the MAE for age groups under 18, 20-50 and over 60 are 1.2, 2.5 and 5.3, respectively

7.2. Age-group prediction

For the VGG age-group detector, we used a quarter of the dataset for validation and the rest for training. A 100-epoch training process took approximately 6 hours to complete. Its validation and training accuracy are respectively 74.3% and 72.95%. Besides, a one-off accuracy, i.e., the prediction of one more or one less than the actual age group, is 92.11% for this model.

This method outperforms some of the previous studies on age prediction in Table 1 when it comes to performance of a caffe-based age-group detector. A high level of accuracy can be achieved through the oversampling method, which provides an average accuracy of 50.27% with MAE 5.13 age groups and an average one-off accuracy of 84.55%. Furthermore, the display of real-time accuracy with predicted age group also allows us to establish the current stability of the result.

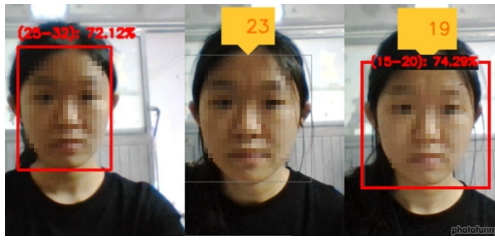


Figure 7: Age-group prediction only (left), Accurate age prediction only (center) and Simultaneous age prediction (right) in single face and standard condition of the 20-year-old girl.

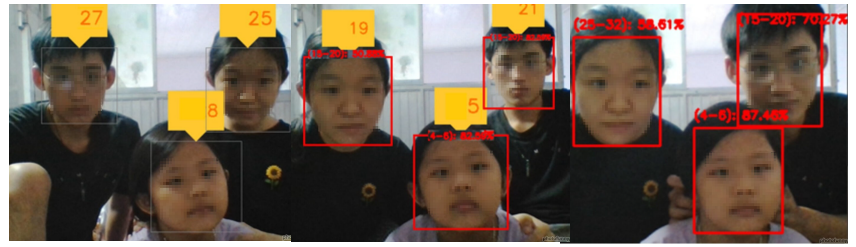


Figure 8: Age-group prediction only (left), Accurate age prediction only (center) and Simultaneous age prediction (right) in multiple faces and unfavorable condition of 3 people whose real age from left to right is 20, 5 and 21.

Table 2: Some Experimental Cases

Age-group test	8-13	8-13	15-20	15-20	25-32	25-32	38-43	38-43	60-	60-
Accurate age test	11	13	16	20	27	30	42	46	65	68
Real age	9,9,7,11, 10	12,12,12, 14, 13	14,16,14, 15,14	20,21,20, 20,20	29,29,30, 25,26	27,27,30, 32,32	37,36,37, 40,38	50,50,47, 51,51	71,71,70, 63,60	76,71,77, 63,72
Condition	Single face, Unfavorable	Single face, Standard	Multiple faces, Standard	Single face, Standard	Multiple faces, Standard	Single face, Unfavorable	Multiple faces, Unfavorable	Multiple faces, Standard	Multiple faces, Standard	Multiple faces, Unfavorable
Age error, accuracy	1.8 83.64%	0.8 93.85%	1.4 91.25%	0.2 99%	2 92.59%	2 93.33%	4.4 90.24%	3.8 91.74%	4.8 92.62%	5.8 91.47%

Table 3: Confusion Matrix of Caffe Model

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-
0-2	0.70	0.26	0.03	0.00	0.01	0.00	0.00	0.00
4-6	0.15	0.57	0.22	0.02	0.03	0.01	0.00	0.00
8-13	0.03	0.17	0.55	0.08	0.14	0.02	0.00	0.01
15-20	0.01	0.02	0.15	0.24	0.51	0.06	0.01	0.01
25-32	0.01	0.01	0.09	0.11	0.61	0.15	0.02	0.01
38-43	0.01	0.01	0.07	0.06	0.46	0.29	0.06	0.05
48-53	0.01	0.01	0.06	0.05	0.26	0.34	0.15	0.13
60-	0.01	0.01	0.06	0.03	0.11	0.27	0.17	0.36

Table 4: Confusion Matrix of VGG Model

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-
0-2	0.86	0.14	0.00	0.00	0.00	0.00	0.00	0.00
4-6	0.40	0.83	0.10	0.00	0.02	0.01	0.00	0.00
8-13	0.00	0.05	0.76	0.04	0.13	0.01	0.00	0.00
15-20	0.00	0.01	0.07	0.33	0.53	0.05	0.01	0.00
25-32	0.00	0.00	0.01	0.02	0.79	0.17	0.01	0.00
38-43	0.00	0.01	0.01	0.00	0.27	0.64	0.06	0.01
48-53	0.00	0.01	0.01	0.01	0.07	0.39	0.40	0.11
60-	0.00	0.00	0.01	0.00	0.02	0.24	0.30	0.43

We attribute the performance of our selective approach to the mutual support of both models. To be more specific, the performance of the caffe-based model for children under 13 is just above average, while VGG-based results are more reliable. But in return, due to confusion with other age groups, the over 60 years old perform less well in terms of efficiency. Moreover, the confusion matrixes in Tables 3 and 4 show that the caffe-based age predictor is heavily biased toward the 25-32 age group. According to the two shown tables, the VGG-Face model estimates age-group labels with greater accuracy than the other, however, the confusion rate (i.e. the difference between the main age label and its adjacent groups) of the caffe-model is smaller than that of the VGG-model in most age groups.

7.3. Comparison between two models

As a result, we find that our age prediction model is slightly influenced by external factors, e.g., real-time lighting conditions and the number of faces within the camera range especially when the boundary overlap of faces occurs. Since our free dataset contains a high number of noisy images, the quality of our prediction fluctuates between some adjacent results. In Table 1, we found that our ensemble method performs better in classification than other methods based on deep neural networks trained on public datasets such as MORPH [14-16] by focusing on the MAE index and the number of divided classes. Furthermore, our method achieved much better results than most deep neural network approaches [15-17], CNN weights are used as feature extractors, while ensemble learning algorithms and majority vote classifiers are applied to generate the most reliable results.

8. Conclusion

A technique using pre-trained fine-tuned VGG-Face weights and Caffe-based to predict accurate age and age-group was described and developed in this paper. The built model showed satisfying results with an acceptable level of reliability, besides the speed and quality of the performances. Although under some unfavorable conditions the accuracy was not as expected, the model presented promising achievement and indicated the orientation to develop and optimize the algorithm in particular and the whole model in general. The contribution of this paper is to demonstrate a combination of methods that are not overcomplicated but still highly effective. From there we want to show that in addition to investing completely in the dataset, we can take advantage of the strengths of the existing well-performed ones and devote our time to develop and optimize the algorithm.

We intend to put under scrutiny other pre-trained weights as well as datasets and thereby create more ethnically balanced datasets which will be very helpful for more efficient feature extraction in order to evaluate and select the most appropriate model with the best performance. Additionally, a completed model for information collection like face recognition can be developed, is a challenge that authors of this work plan to overcome.

Conflict of Interest

All authors have no conflicts of interest to disclose.

Acknowledgment

This research was supported in terms of equipment and expertise by the Faculty of Electrical and Electronic Engineering, Ho Chi Minh City University of Technology (HCMUT), Vietnam National University, Ho Chi Minh City (VNU-HCM).

References

- [1] A. Dantcheva, P. Elia and A. Ross, "What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics," *IEEE Transactions on Information Forensics and Security*, **11**(3), 441-467, 2016, doi: 10.1109/TIFS.2015.2480381.
- [2] Q. Zhang, D. Zhou and X. Zeng, "HeartID: A Multiresolution Convolutional Neural Network for ECG-Based Biometric Human Identification in Smart Health Applications," *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.2707460.
- [3] R. D. Labati, A. Genovese, V. Piuri and F. Scotti, "Weight estimation from frame sequences using computational intelligence techniques," *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, 2012, doi: 10.1109/CIMSA.2012.6269603.
- [4] E. Eiding, R. Enbar and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," *IEEE Transactions on Information Forensics and Security*, **9**(12), 2170-2179, 2014, doi: 10.1109/TIFS.2014.2359646.
- [5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, doi: 10.1109/CVPR.2001.990517.
- [6] T. Surasak, I. Takahiro, C. Cheng, C. Wang and P. Sheng, "Histogram of oriented gradients for human detection in video," *5th International Conference on Business and Industrial Research (ICBIR)*, 2018, doi: 10.1109/ICBIR.2018.8391187.
- [7] P. Kruachottikul, N. Cooharajanone, G. Phanomchoeng, T. Chavarnakul, K. Kovitangoon and D. Trakulwanont, "Deep learning-based visual defect-inspection system for reinforced concrete bridge substructure: a case of Thailand's department of highways" *J Civil Struct Health Monit*, 2021, doi: 10.1007/s13349-021-00490-z
- [8] P. Kruachottikul, N. Cooharajanone, G. Phanomchoeng, T. Chavarnakul, K. Kovitangoon, D. Trakulwanont, and K. Atcharyachanvanich, "Bridge Sub Structure Defect Inspection Assistance by using Deep Learning," *IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, 2019, doi: 10.1109/ICAwST.2019.8923507.
- [9] G. Guo, Y. Fu, C. R. Dyer and T. S. Huang, "Image-Based Human Age Estimation via Manifold Learning and Locally Adjusted Robust Regression," *IEEE Transactions on Image Processing*, **17**(7), 1178-1188, 2008, doi: 10.1109/TIP.2008.924280.
- [10] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," *CVPR*, 2011, doi: 10.1109/CVPR.2011.5995404.
- [11] X. Geng, C. Yin and Z. -H. Zhou, "Facial Age Estimation by Learning from Label Distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(10), 2401-2412, 2013, doi: 10.1109/TPAMI.2013.51.
- [12] F. Li, "Linear classification: Support Vector Machine, Softmax," 2021. [Online]. Available: <https://cs231n.github.io/linear-classify/>. [Accessed 9 June 2021].
- [13] I. Huerta, C. Fernández, A. Prati, "Facial Age Estimation Through the Fusion of Texture and Local Appearance Descriptors," *Computer Vision - ECCV 2014 Workshops*, doi: 10.1007/978-3-319-16181-5_51
- [14] A. Lanitis, C. J. Taylor and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(4), 442-455, April 2002, doi: 10.1109/34.993553.
- [15] H. Takimoto, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "Robust gender and age estimation under varying facial pose." *Electronics and Communications*, 2018, doi: 10.1002/ecj.10125
- [16] 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*, **29**(12), 2234-2240, 2007, doi: 10.1109/TPAMI.2007.70733.
- [17] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," *CVPR*, 2011, doi: 10.1109/CVPR.2011.5995404.
- [18] Z. Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong and S. Yan, "Facial Age Estimation With Age Difference," *IEEE Transactions on Image Processing*, **26**(7), 3087-3097, 2017, doi: 10.1109/TIP.2016.2633868.
- [19] A. Kumar, "Keras – Categorical Cross Entropy Loss Function," 2020. [Online]. Available: <https://vitalflux.com/keras-categorical-cross-entropy-loss-function/>. [Accessed 2 June 2021].
- [20] Gilbert Tanner, "Introduction to Deep Learning with Keras," 2019. [Online]. Available: <https://towardsdatascience.com/introduction-to-deep-learning-with-keras-17c09e4f0eb2>. [Accessed 10 June 2021].
- [21] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, doi: 10.1109/FGR.2006.78.
- [22] "Pandas DataFrame," 2019. [Online]. Available: <https://www.geeksforgeeks.org/python-pandas-dataframe/>. [Accessed 2 June 2021].
- [23] F.-F. Li, "Linear classification: Support Vector Machine, Softmax,," 2021. [Online]. Available: <https://cs231n.github.io/linear-classify/>. [Accessed 9 June 2021].
- [24] James, Gareth Michael, "Majority vote classifiers: theory and applications," *Stanford University*, 1998.