IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-6, Special Issue-7, Sept. 2018 E-ISSN: 2347-2693

Comparative Evaluation on Supervised Learning Based Age Estimation

A. Annie Micheal[1*] , P. Geetha[2] , A. Saranya[3]

[1,2,3] Dept. of Information Science and Technology, Anna University, Chennai, India

[*]*Corresponding Author: annymick@gmail.com*

Available online at: www.ijcseonline.org

***Abstract*—** Facial age estimation has got more consideration in the area of computer vision for the past few years. Age estimation is a troublesome task since the distinction between facial pictures with age variations is difficult. In this work, we analyze the problem of age prediction by means of SVR Model and deep learning technique. This paper attempts to find out the efficiency of SVR and Convolution neural network (CNN) on age estimation. Local features such as wrinkles and texture are extracted using Gabor filter, Local Binary Pattern (LBP) and Local Phase Quantization (LPQ). The three features are combined together and the dimension of the feature vector is reduced using Principle Component Analysis. Support Vector Regression (SVR) is utilized to predict the age of an individual. In CNN, the datasets are fine-tuned utilizing the pre-trained VGG-16 model which can group pictures into 1000 categories. The experimental results on the IMDB-WIKI dataset, the ICCV datasets and MORPH 2 dataset shows that CNN outperforms the local feature based SVR model in predicting the age.

Keywords: Convolutional Neural Network, Local binary Pattern, Local Phase Quantization, Gabor Filter, Support Vector Regression.

I. INTRODUCTION

Automatic age estimation task aims to estimate a person's age based on features extracted from face image. Age estimation has encountered many problems such as different poses, inter-person and lighting variation, face orientation and occlusions in image. The complex task in age estimation is to manually design a suitable feature learning method. Estimating human age automatically via facial image analysis has lots of potential real-world applications such as Internet access control, underage cigarette-vending machine use, age-based retrieval of face images, age prediction systems for finding lost children, advanced video surveillance, demographics statistic collection, human computer interaction, targeted advertising. Estimating human age from a facial image requires a great amount of information from the input image. Extraction of these features is important since the performance of an age estimation system will heavily rely on the quality of extracted features. Lots of research on age estimation has been conducted towards aging feature extraction. In our work, we evaluate the performance of age estimation using local feature extraction methods and CNN. Wrinkle features are extracted using Gabor filters. Local Binary Pattern and Local Phase Quantization are utilized to extract texture features. Principle Component Analysis is utilized for dimension reduction and age of the person is estimated using Support Vector Regression. The estimated age using the local feature extraction method is compared with the age estimated using CNN. CNN claim to have the best performances in age estimation.

The remaining of this manuscript is structured as follows. In Section. 2 some related works are reviewed. In Section. 3, presents the proposed approach. Section. 4 describes the experimental results and analysis. Section. 5 includes the conclusion of the paper.

II. RELATED WORK

There are numerous explores established for age estimation. In the most recent years, Bio-Inspired Features (BIF) [1] have been have been reliably utilized for age estimation [2, 3]. These feed-forward models comprises of various layers entwining convolution and pooling processes. Initially, an input picture is mapped to a higher-dimensional space by convoluting it with a bank of multi-scale and multi-orientation Gabor filters. Afterwards, a pooling step down scales the outcomes with a non- linear reduction, typically a MAX or STD operation, dynamically encoding the outcomes into a vector signature. In [4], Wen-Bing Horng, Cheng-Ping Lee and Chun-Wen Chen extracted the facial wrinkles using the Sobel edge operator. The facial pictures were grouped into 4 age groups using two back propagation network. The distinguishing proof rate of the framework accomplishes 81.58%. Suo et al. [5] implemented a face model [6] to represent the extensive varieties of facial structures, and increase it with age and hair features. They additionally embrace a dynamic Markov process on this graph representation for face aging. Facial wrinkles pattern were studied by Hayashi et al. [7] for estimating age, from a database of controlled face images taken from 300 subjects of

Japanese people ranging from 15 to 64 years of age. Skin regions were extracted for the purpose of quantifying and distinguishing the types of wrinkles, short or long. The wrinkles were extracted using Digital Template Hough Transform and enhanced by histogram equalization on the raw skin regions. At last, a look-up table based on the number and type of wrinkles was utilized to characterize the subject's age. The work exhibit that wrinkle features alone is not a good indicator of age. Ricanek et al. [8] used Active Appearance Model to extract the shape features from a facial image with 161 landmarks. Different methodologies are there to partition the face images into numerous sub-regions, and features are extracted from these regions, afterward consolidate them together. Horng, Lee and Chen [9] classified the age into four groups: Infants Young, Middle- aged adults and Senior. Back Propagation Networks was utilized for arranging the face age, in which, one was utilized to guarantee whether the face picture was a child and the other one was to group whether the face picture was a youthful grown-up, moderately aged grown-up or senior grown-up. Lanitis et al. [10] estimated the work of distinct classifiers for age estimation, containing a quadratic function classifier, the nearest neighbor classifier and the Artificial Neural Networks (ANNs). The experiment was performed on a small database containing 400 images ranging from 0 to 35 years. In [11], Kwon and Lobo estimated the age using wrinkle features and used six ratios of distances between eyes, noses, mouth, chin and forehead. The age was categorized into three groups: infants, young, and senior adults. C. Taylor et al. [12] utilized Active Appearance Models (AAM) to extract the shape and appearance information from the facial images and they joined together to obtain a parametric description. Then an aging function is built based on the relationship between the parametric description and the age of the individual. Age estimation is still a challenging issue regarding accuracy.

III. PROPOSED APPROACH

A. AGE ESTIMATION USING CNN

The proposed methodology is shown in Fig. 1. Viola-Jones algorithm is used to detect the face. The initial step of creating and training a CNN is to define the network layer which is based on the VGG-16 model. The number of layers included is dependent on the particular application of data. Feature representation and age estimation are the two key modules of facial age estimation. In our proposed work, the pre-trained VGG face model is used as the basic model for age estimation. The pre-trained VGG face model is fine-tuned with three datasets to train the CNN model with feature representations. The datasets used for fine-tuning are IMDB, ICCV and MORPH 2 age datasets. The datasets are fine-tuned in different deep networks (CNN) to get the new model. Finally, the testing images are compared with the trained deep model. The distribution got by deep model provides information to predict the age.

a. VGG-16

VGG-16 is a pre-prepared model in CNN. This model is trained on a subset of the ImageNet database which is utilized as a part of the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). VGG-16 is prepared on in excess of a million pictures and can classify images into 1000 object classes.

b. FINE TUNING

Fine-tuning is the process to take a network model that has already been trained for a given task and make it perform a second similar task. The original task would be classifying the images into 1000s of categories. If the new task is to classify the images into smaller number of categories then the last layer (fully connected layer) of network is removed while fine-tuning and trainable is set to false for all the other layers as they have already been trained.

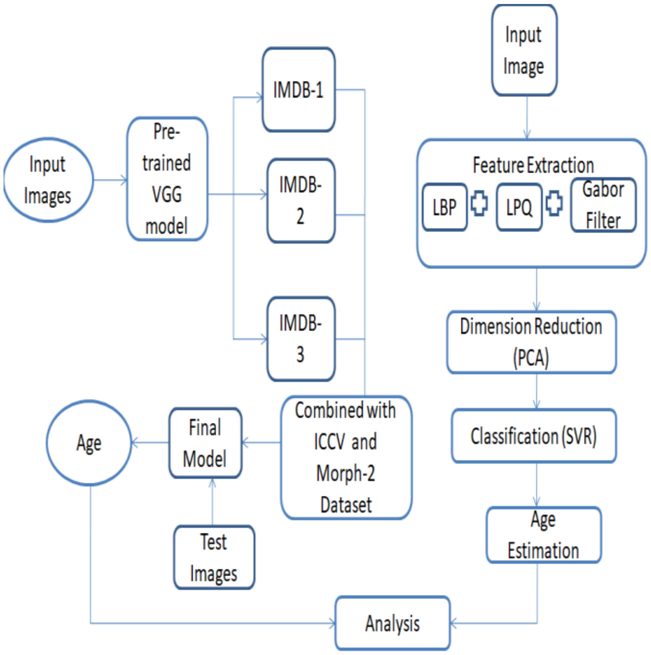


Fig. 1. System Architecture of Proposed Approach

c. CONVOLUTIONAL NEURAL NETWORK

CNN is a feed-forward network that can extract topological properties from an image and it is effective in areas such as image recognition and classification. CNN is comprised of one or more convolution layers and then followed by one or more fully connected layer. The architecture of CNN is designed to take advantage of the 2D structure of an input image. This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. The four layers of convolutional neural network are Convolution layer, Non Linearity (ReLU), Pooling or Sub sampling Layer and Fully Connected Layer

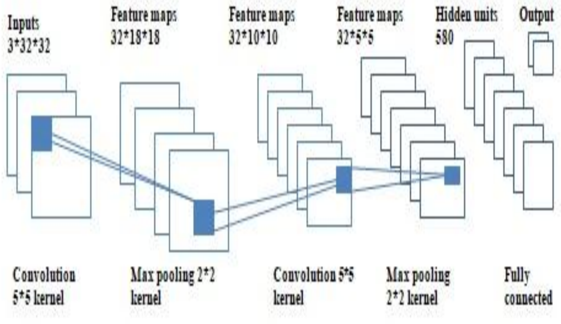
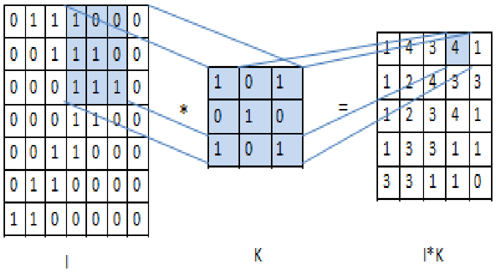


Fig. 2. CNN Architecture

1. CONVOLUTION LAYER

The principal layer of CNN is convolution layer which convolve the filter with the picture. It convolve (slide) over every spatial area of the image. Each node has similar weights in a layer. Every node has its common weight convolution registered on an open field somewhat moved from that of its neighbor. Two fundamental parameters of this layer are stride and padding. Stride controls how the channel convolves around the information volume and padding controls the volume by applying zero padding.



0	1	1	1	0	0	0
0	0	1	1	1	0	0
0	0	0	1	1	1	0
0	0	0	1	1	0	0
0	0	1	1	0	0	0
0	1	1	0	0	0	0
1	1	0	0	0	0	0

1	0	1
0	1	0
1	0	1

1	4	3	4	1
1	2	4	3	3
1	2	3	4	1
1	3	3	1	1
3	3	1	1	0

Fig.3. Convolutional Layer

2. RECTIFIED LINEAR UNIT (RELU) LAYER

ReLU layer is a tradition to apply a non-linear layer. It is also called as activation layer. The reason for this layer is to acquaint non-linearity with a framework that fundamentally has quite recently been processing direct tasks amid the convolution layers. ReLU work far better because the network is able to train a lot faster without making a significant difference to the accuracy. This layer changes all the negative activations to 0 and builds the non-linear properties of the model.

3. POOLING LAYER

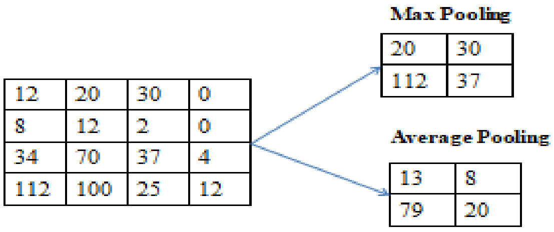
Pooling layer is otherwise known as the down sampling layer. This layer has two categories such as max pooling and average pooling. The max pooling basically takes the 2*2 size filter and a stride of same length. It then applies it to the input volume and outputs the maximum number in every sub region that the filter convolves around. Similarly the average pooling output will be the average of the sub region weights. The two types of pooling layer are Maximum Pooling Layer and Average Pooling Layer

I. MAXIMUM POOLING LAYER

Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. It returns the maximum value of the rectangular region of its output.

II. AVERAGE POOLING LAYER

An average pooling layer outputs the average values of rectangular regions of its input. The Pool Size property determines the size of the rectangular regions. For example, if Pool Size is (2,3) , then the layer outputs the average value of regions of height 2 and width 3.



12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

Max Pooling

20	30
112	37

Average Pooling

13	8
79	20

Fig. 4. Pooling Layer

4. FULLY CONNECTED LAYER

Fully connected layer is the last layer of CNN. This layer basically takes an input volume and outputs an N dimensional vector where N is the number of classes that the program has to choose from. Fully connected layers connect every neuron in one layer to every neuron in another layer. After several convolution and max pooling layers, the high-level reasoning in the neural network is done via fully connected layers.

B. AGE ESTIMATION USING LOCAL FEATURES

The LBP, LPQ and Gabor wavelets are used to extract texture and wrinkles. The two important steps of age prediction system are feature extraction and classification. Once the features are extracted and concatenated successfully, the dimensions of the feature vector are reduced using Principle Component Analysis (PCA). SVR is used to predict the age of a person.

a. LOCAL BINARY PATTERN

The LBP operator was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. LBP operates in a local circular region by taking the difference of the central pixel with respect to its neighbors. It is defined as

$$LBP_{R,N} = \sum_{n=0}^{n-1} a(b_n - b_c).2^n , \qquad (1)$$

a(bn-bc)=1, if *bn ≥bc* otherwise 0. Where *b*n and *b*c are the gray values of the central pixel and its neighbor, respectively, n is the index of the neighbor, *R* is the radius of the circular neighborhood and *N* is the number of the neighbors.

b. LOCAL PHASE QUANTIZATION

Local phase quantization operator is an image descriptor based on the blur invariance property of the Fourier phase spectrum. LPQ is a blur insensitive texture descriptor based on the blur invariance property of the Fourier phase spectrum. In this method LPQ codes are computed in local image windows using discrete Fourier transform (DFT) and the results are presented as a histogram. Let *Kx* defines the N-by-N neighborhood around the pixel position x of the image f(x). The two dimensional (2-D) DFT or, more precisely, short term Fourier transform (STFT) of this image window is defined by

$$F(c,d)= \sum_{y \in K_x} f(x-y) e^{-j2\pi c^E y} = z_c^E f_x , \quad (2)$$

where, *zc* is the basis vector of the 2-D DFT at frequency c, and *fx* is the vector containing all N^2 pixels in *Kx*. Using 2-D convolutions f(x) $e^{-j2\pi c^E x}$ for all c is the efficient way of STFT implementation. As the basis functions are separable, this 2-D computation can be performed by using 1-D convolutions for the rows and columns successively. Only the complex coefficients of c1 = $[v, 0]^E$, c2 = $[0, v]^E$, c3 = $[v, v]^E$, c4 = $[v,-v]^E$ (v is a scalar frequency) are considered in LPQ. For each pixel position this results in a vector given by

$$F_X^G = [F(c_1, x), F(c_2, x), F(c_3, x), F(c_4, x)], \text{ (3) and}$$

$$F_x = [\mathrm{Re}\{F_x^G\}, \mathrm{Im}\{F_x^G\}]^E , \qquad (4)$$

where Re{.} and Im{.} return real and imaginary parts of a complex number, respectively. The corresponding 8-by-N^2 transformation matrix is

$$Z = [\mathrm{Re}\{z_{c1}, z_{c2}, z_{c3}, z_{c4}\}, \mathrm{Im}\{z_{c1}, z_{c2}, z_{c3}, z_{c4}\}]^E \quad (5)$$

So that *Fx=Zfx*. Next, *Mx* is computed for all images positions i.e., $x \in \{x_1, x_2, ..., x_K\}$, and the resulting vectors are quantized using simple scalar quantizer

uj= 1 , if mj ≥0 otherwise 0 (6)

Where *mj* is the *jth* component of *Mx.* The quantized coefficient are represented as integer values between 0 and 255 using binary coding

$$o = \sum_{j=1}^{8} u_j 2^{j-1}$$

(7)

Finally, the histogram of these integer values is used as a feature vector.

c. GABOR FILTER

The neighborhood highlights identified with wrinkles are separated from an arrangement of Gabor wavelets for being powerful to clamor, for example, hair, facial hair, and shadows. Gabor channel are equipped for deciding the introduction of a wrinkle and is utilized for the wrinkle highlight extraction. By outlining a wrinkle particular Gabor channel, a commotion hearty wrinkle highlight can be removed. An arrangement of Gabor wavelets is the result of a Gaussian envelope and a plane wave. The Gabor change at a specific position of a picture can be processed by a convolution with the wavelets. All highlights get from the size of the subsequent complex picture.

d. PRINCIPAL COMPONENT ANALYSIS

PCA is a factual way to deal with discovers the key highlights of a circulated dataset in view of the aggregate difference. Given an arrangement of multivariate conveyed information in X-Y organize framework, PCA first finds the most extreme varieties of the first datasets. These information focuses are then anticipated onto another hub called U-V arrange framework. The heading of U and V axis is known as main parts. All in all, the picture measurement diminishment strategy by PCA can be united into four noteworthy advances: picture standardization, finding the covariance grid of the picture information, ascertain the Eigen vectors and Eigen estimations of the covariance framework and ultimately changing picture information into new premise. Standardization is the way toward subtracting the picture with their mean esteem can enhance the signal-to-noise ratio (SNR) and Discrimination Power (DP) of the picture. Covariance matrix is figured to find the best change estimation of a picture. Eigen vector with the biggest Eigen esteem is the heading of most prominent variety in the picture. The Eigen vector lattice speaks to the central highlights of the picture information which is utilized to changes the first picture dataset into another axis with lessened measurement.

e. SUPPORT VECTOR REGRESSION

Support Vector Machine can likewise be utilized as a regression technique, keeping up all the fundamental highlights that describe the calculation (maximal margin). The Support Vector Regression (SVR) utilizes an indistinguishable standard from the SVM for grouping, with just a couple of minor contrasts. On account of relapse, a margin of resilience (epsilon) is set in guess to the SVM which would have effectively asked for from the issue. The principle thought is to limit mistake, individualizing the hyper plane which amplifies the edge, remembering that part of the error is endured.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The different facial dataset which are used to fine-tune on the pre-trained model are IMDB-WIKI, ICCV and MORPH 2 dataset. IMDB-WIKI dataset is the largest publicly available dataset of facial images with gender and age label with 62,329 images and ICCV 2015 ChaLearn Looking at People workshop dataset which is a small dataset contains 7,224 images with apparent ages and standard deviation. Morph dataset which is a large dataset contains 55,608 facial images with about three aging images per person ranging from16 to 77 years old. The CNN model is trained using 18,424 images from the IMDB, ICCV and Morph-2 datasets. The input images for the Local Feature approach are taken from all the above mentioned datasets. Initially the input image is converted to the gray scale image. From the above mentioned datasets, 1600 images are taken and 75% is used for training the model and 25% is used for testing the model.

Table 1. Mean Absolute Error for CNN model and Local Feature Approach using IMDB, ICCV and MORPH datasets.

Methods	IMDB	ICCV	MORPH 2
LBP + LPQ + Gabor	4.62	4.83	4.28
CNN	1.016	1.382	1.831

Mean Absolute Error is the estimation the accuracy by calculating the age differences using predicted and ground truth ages. Table.1 shows the accuracy results using IMDB, ICCV and Morph datasets for both the CNN and Local feature approaches. The accuracy of CNN approach results with a maximum age difference of 1.831. The maximum age difference of SVR based local approach is 4.83. Hence CNN approach has more accuracy with less age differences when compared to the SVR approach.

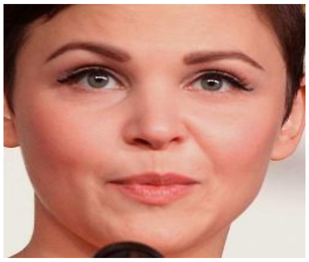


Fig.6. Input Image

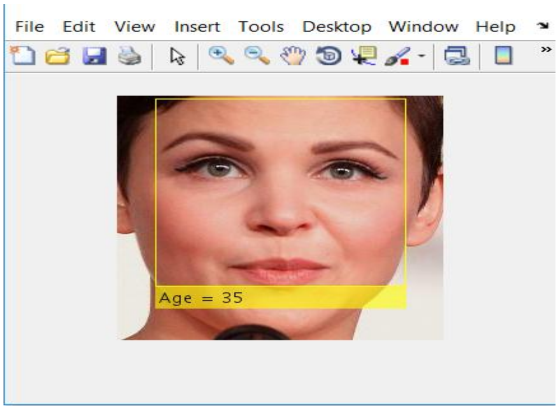


Fig.7. Output Image

V. CONCLUSION

In our proposed work, Viola-Jones algorithm is used to detect the frontal faces from the input images and VGG face model is used as the basic model. Using the pre-trained model, different datasets are fine-tuned. The predicted age distributions are extracted from the fine-tuned CNN model. Finally the trained CNN model is used to get the result by comparing both the training and testing images. The accuracy of the age estimation system is calculated based on the Mean Absolute Error. The results of age estimation using CNN is compared with the results local feature age estimation system. The proposed local feature age estimation system involves various local feature extraction techniques such as LBP, LPQ and gabor filter to extract wrinkle and texture features. PCA is used to reduce the dimension of the feature vectors and given as the input to the SVR classifier with the ground truth ages to predict the age. The results obtained demonstrate that the proposed descriptor CNN outperforms the local feature extraction approaches. This proposed work implements age estimation using frontal face images. In future, the work can be extended to detect faces in multi-pose and unconstrained environment. Some additional features can also be included in future to improve the accuracy of the age estimation system.

REFERENCES

[1]. Maximilian, Riesen huber and Tomaso Poggio (1999),"Hierarchical models of object recognition in cortex", *Nature neuroscience*, vol. 2, no. 11, pp. 1019–1025.

[2]. XinGeng, Chao Yin, and Zhi-Hua Zhou (2013), ‘’Facial age estimation by learning from label distributions’’, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 35,no.10, pp. 2401–2412.

[3]. Hu Han, Charles Otto, and Anil K Jain (2013)," Age estimation from face images: Human vs. machine performance", *In International Conference on Biometrics (ICB)*. IEEE, pp.1-8.

[4]. Wen-Bing Horng, Cheng-Ping Lee and Chun-Wen Chen (2001),"Classification of Age Groups Based on Facial Features",

Tamkang Journal of Science and Engineering, vol.4, no.3, pp.183-192.

[5]. J. Suo, Min Feng, S. Zhu, S. Shan, X. Chen (2007), "A multi-resolution dynamic model for face aging simulation*", Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp. 1-8.

[6]. Z. J. Xu, H.Chen, and S. C. Zhu (2005),"A high resolution grammatical model for face representation and sketching". *IEEE CVPR*, pp: 470-477.

[7]. J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu (2002), ''Age and Gender Estimation from Facial Image Processing'', *the41st SICE Annual Conference*, vol. 1, pp. 13 -18, Aug.

[8]. K. Ricanek , Y. Wang , C. Chen , S. J. Simmons (2009) "Generalized multi-ethnic face age-estimation", *in Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on. IEEE*, pp. 1-6.

[9]. Wen – bingHorng, cheng-Ping Lee and chun-Wen chen (2001),"classification of age groups based on facial feature", *Journal of Science and Engineering*, vol. 4, no.3, pp.183-192.

[10]. Y. Kwon and N. Da Vitoria Lobo (1999), "Age classification from facial images" *Computer vision and image understanding*, vol. 74, no. 1, pp.1–21.

[11]. A. Lanitis, C. Draganova, and C. Christodoulou (2004), "Comparing different classifiers for automatic age estimation". *In Proceedings of. IEEE Transactions on SMC-B*, vol. 34, no.1, pp. 621-628.

[12]. A. Lanitis, C. Taylor, T. Cootes (2002), "Toward Automatic Simulation of Aging Effects on Face Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455.