International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 04, April-2020

Gender Recognition and Age Approximation using Deep Learning Techniques

Shubham Patil*, Bhagyashree Patil*, Ganesh Tatkare*
Guide: Prof. Saumya Salian*
*Department of Computer Engineering,
*Datta Meghe College of Engineering
Mumbai University

***Abstract*--- Age and gender that are the two key facial attributes, play a foundational role in social interactions, making age and gender estimation from one face image a crucial task in intelligent applications, like access control, human-computer interaction, enforcement, marketing intelligence and visual surveillance. The basic aim of this paper is to develop an algorithm that estimates age and gender of a person correctly. One of the most widely used techniques is haar cascade. In this paper we propose a model which can predict the gender of a person with the assistance of Haar Cascade. The model trained the classifier with different male and female images as positive and negative images. Different facial features are extracted. With the assistance of Haar Cascade classifier will determine whether the input image is male or female. We made use of Deep-Convolution neural network. It works efficiently even with limited data. For the age approximation task, the paper makes use of caffedeep learning framework. Caffe provides expressive architecture, extensible code. Caffe can process over 60M photos per day. This makes it one of the fastest convent implementation available.**

Keywords--- Gender recognition, Age classification, Haar cascade, Caffe deep learning framework.

I. INTRODUCTION

Human face may be a storehouse of various information about personal characteristics, including identity, emotional expression, gender, age, etc. the looks of face is affected considerably by aging. This plays a significant role in non-verbal communication between humans. Age and gender, two key facial attributes, play a really foundational role in social interactions, making age and gender estimation from one face image a very important task in machine learning applications, like access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance.

Automatic gender classification and age detection may be a fundamental task in computer vision, which has recently attracted immense attention. It plays a very important role in an exceedingly wide selection of the real-world applications like targeted advertisement, forensic science, visual surveillance, content-based searching, human-computer interaction systems, etc. for instance we are able to use this

method to display advertisement supported different gender and different age bracket. This method may be employed in different mobile applications where there's some age restricted content in order that only appropriate user can see this content. However, gender classification and age approximation is still a difficult task. We propose a model which can first perform feature extraction on the input image which can classify eyes, lips, beard, hair, etc. Supported these features the model will classify the gender as male or female. We've used Haar Cascade for feature extraction purpose. Age is estimated with the assistance of Caffe Model. The age classifier takes an image of an individual's face of size 256x256 as an input to the algorithm that's then cropped to 227x227. The age classifier returns a integer representing the age range of the individual. There are 8 possible age ranges , that the age classifier returns an integer between 0 and seven. The gender classifier returns a binary result where 1 indicates male and 0 represents female.

II. FEATURE EXTRACTION

A. Gender classification

Images may not be perfect. There are many noises which are redundant. This can decrease system performance. To increase accuracy rate we have to make proper and effective feature extraction. This can be global or local which depends on shape, color, orientation.

1) Edge detection:Edge feature is mostly used for detecting the object. It finds the discontinuities in gray level. We can say that edge is the boundary between the regions.

2) Haar– like features: [14]Viola and Jones proposed an algorithm which is called Haar-Classifiers for rapid object detection and pedestrian detection is applied. it is done with the haar like features which can be calculated efficiently by using Adaboost classifier and integral images in cascade classifier. Haar-like features can have high accuracy and in low cost. Haar cascade is mostly used for face detection because of its easy calculation.

3) Detector using haar -Like features: In face detection, the image is first scanned, looking for patterns with indicate the presence of a face in the image. This is done by using haar-like features. The haar like features are created by two or three adjacent rectangles with different contrast values.

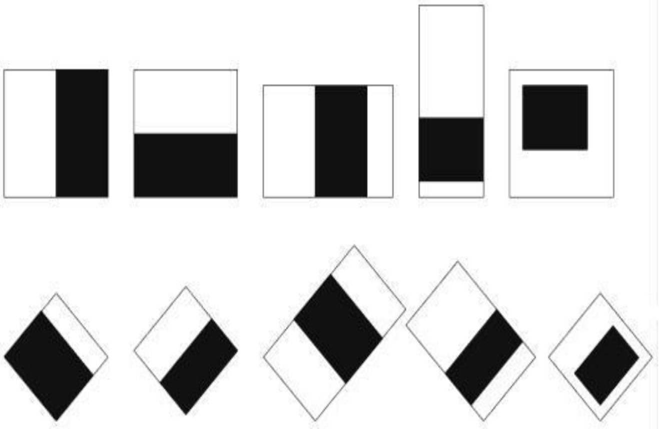


Fig 1. HAAR FEATURES

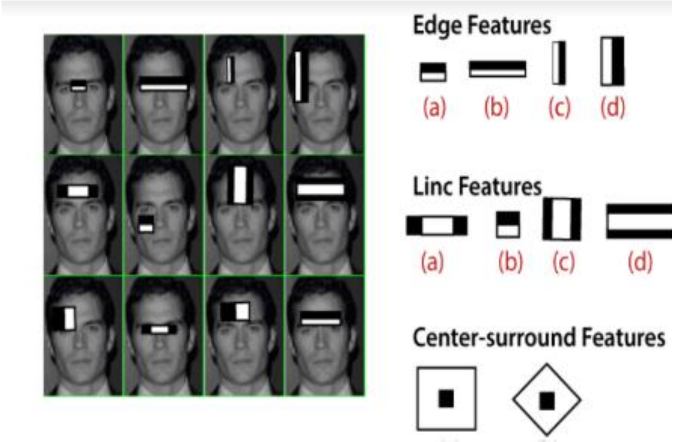


Fig 2. WORKING OF HAAR FEATURES

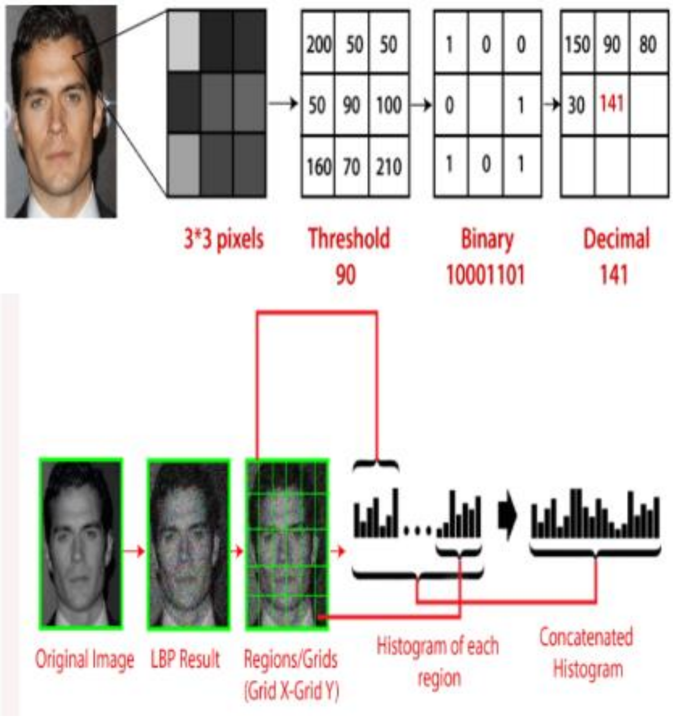
There are separate intensity values of black and white pixels which gives dark and light regions. Any object can be detected by using haar like features. We just have to adjust the size of rectangle that we can scale up and down the image.

4)Texture feature: This feature was proposed by [15]Robert M. Haralick in 1973. These features make use of the statistics which summarize the relative frequency distribution which describes how one gray tone is spatially related to the graytone. Local Binary Pattern Algorithm is used for extracting texture features.

This is simple but effective algorithm to extract texture features. Using LBP computation an intermediate image is generated which describes the original image in specific way. For concept of sliding window the parameters like radius and neighbors are used.

The input facial image is grayscale . This approach get the block of this image as 3x3 matrix. The 3x3 matrix contains the intensity of each pixel (0-255). We take the central value of the matrix and use it as threshold.

This value is used to define the new values of the matrix. Each neighbor of the central value is set as a new binary value. The value is set to 1 if it is greater than the threshold value and 0 if the value is lower than the threshold value. The matrix now contains only binary values. We now concatenate each binary value line by line or clockwise but the out will be same. Then convert the binary value into decimal value. Likewise each pixel in the matrix is converted into decimal value. The resultant image represents better characteristics of original image. Histograms are derived from each such matrix of image and all the histograms are concatenated which show the better characteristics of the original image.



200	50	50
50	90	100
160	70	210

1	0	0
0		1
1	0	1

150	90	80
30	141	

Fig 3. HISTOGRAM GENERATION

B. Age Classification

Steps for feature extraction:

1. An input image or class of images.
2. Pre-trained model(.caffe-model)-binary file which has the weights, gradients and biases for each layer of the network.
3. Model definition (.prototxt) file which has the network structure that is used.
4. Target feature extraction layer(eg:-“fc7” in Alexnet)

III. PROPOSED ALGORITHM

Recognition of gender is kind of difficult when the image is captured from far distance by using haar-like features. For this problem we have used simple but effective idea. We applied cascaded method. The paper uses ROI (Region of interest) as our face. We returned the ROI image to classifier. In this paper, we tried to detect the female face. We trained our haar cascade classifier by 500 female and 500 male images. We used frontal face images to train that included external features like hairstyle, makeup, accessories such as earings and glasses. The object that this paper is trying to detect is positive in xml training. In our case we use only female and male as positive and negative image. This means the image which is not female is male

This is very simple but efficient gender recognition problem.

B. Convolution Neural Network

The main building block of CNN is the convolutional layer. Convolution is a mathematical operation to merge two sets of information. In our case the convolution is applied on the input data using a convolution filter to produce a feature map.

There are a lot of terms being used so let’s visualize them one by one.

On the left side is the input to the convolution layer, for example the input image. On the right is the convolution filter, also called the kernel, we will use these terms interchangeably. This is called a 3x3 convolution due to the shape of the filter.

We perform the convolution operation by sliding this filter over the input. At every location, we do element-wise matrix multiplication and sum the result. This sum goes into the feature map. The green area where the convolution operation takes place is called the receptive field. Due to the size of the filter the receptive field is also 3x3.

Caffe for Age approximation:

Caffe is a CNN framework which allows researchers and other practitioners to build a complex neural network and train it without need to write much code. For estimation of age using the convolution neural network, gathering a large dataset for training the algorithm is a tedious and time consuming job. The dataset needs to be well labeled and from social image database which has the private information of the subjects i.e. age

Network Architecture:

The network architecture used for age approximation in based on the paper of [2]G.Levi and T.Hassner. This network is intended to be shallow to prevent over-fitting. All the three colors i.e. Red, Green, Blue are processed directly. The images

are scaled to 256 x 256 and cropped to 227 x 227. The network consists of 3 convolution layers followed by 3 fully connected layers.

Layer 1: Here, filters of size 3x7x7 are convoluted with stride 4 and padding 0, which results in output of size 96x56x56 which is followed by maximum pooling which reduce the size and local response normalization.

Layer 2: 256 filters sized 96x5x5 are convoluted with stride 1 and padding 2, that results in output of size 256x28x28. This is also followed by maximum pooling and LRN reducing the size to 256x14x14.

Layer 3: 256 filters of size 256x3x3 are concoluted with stride 1 and padding 1.

The fully connected layers:

1. The first fully connected layer which gets the results from the last convolution layer and contains 512 neurons, followed by ReLU and dropout layer.
2. The second fully connected layers gets the output from the previous layer of 512 dimension and again contains 512 neurons, followed by ReLU and dropout layer.
3. The last fully connected layer maps to the final classes for age.

The output of the fully connected layers is fed to soft-max layer that assigns probability for each class. The given test image is tested with maximum probability and the prediction is made.

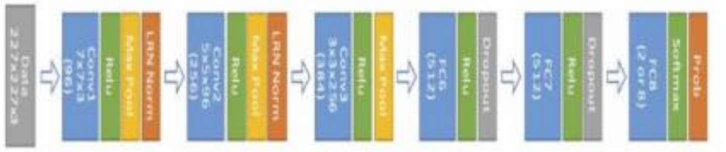
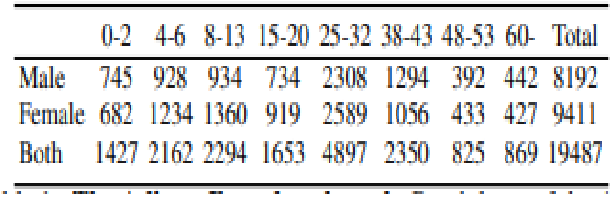


Fig 4. CAFFE NETWORK ACRCHITECTURE

The Adience Dataset:

The benchmark for this problem, introduced by Eran Eidinger et al. [3], uses the Adience dataset which is composed of images scraped from Flickr.com albums that were labeled for age and gender. The benchmark uses 8 classes for age groups (0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, 60+), and therefore we treated both gender prediction and age prediction as a classification problem. The Adience dataset is relatively small (containing 34,795 images), so we also used the IMDB+Wiki dataset which is the largest dataset publicly available for age and gender (containing 523,051 face images).



	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

Fig 5. ADIENCE DATASET DISTRIBUTION

Technical Details:

Local Response Normalization (LRN).:

After the primary 2 pooling layers, there are local response normalization (LRN) layers. LRN could be a technique that was first introduced in as the way to assist the generalization of deep CNNs. The idea behind it's to introduce lateral inhibition between the various filters in a very given convolution by making them “compete” for big activations over a given segment of their input. This effectively prevents repeated recording of the identical information in slightly different forms between various kernels watching the identical input area and instead encourages fewer, more prominent, activations in some for a given area. If a(x,y) is that the activation of a neuron by applying kernel i at position (x, y), then it’s local response normalized activation b(x,y) is given by

$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=max(0,i-n/2)}^{min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^\beta$$

where k,n,α, and β are all hyper-parameters. The parameter n is that the number of “adjacent” kernel maps (filters) over which the LRN is run, and N is that the total number of kernels therein given layer.

Softmax:

At the highest of the proposed architecture lies a softmax layer, which computes the loss term that's optimized during training and also the category probabilities during a classification. While some loss layers like multiclass SVM loss treat the output of the ultimate fully connected layer because the class scores, softmax (also called multinomial logistic regression) treats these scores because the unnormalized log probabilities of the classes. That is, if we've got zi is the score assigned to class i after the ultimate fully connected layer, then the softmax function is

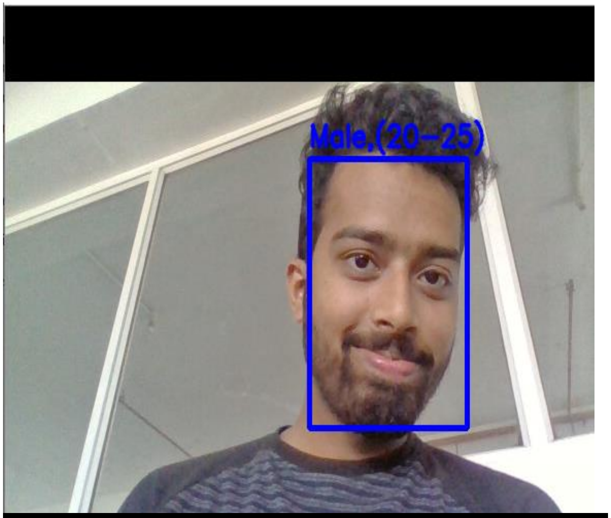
$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

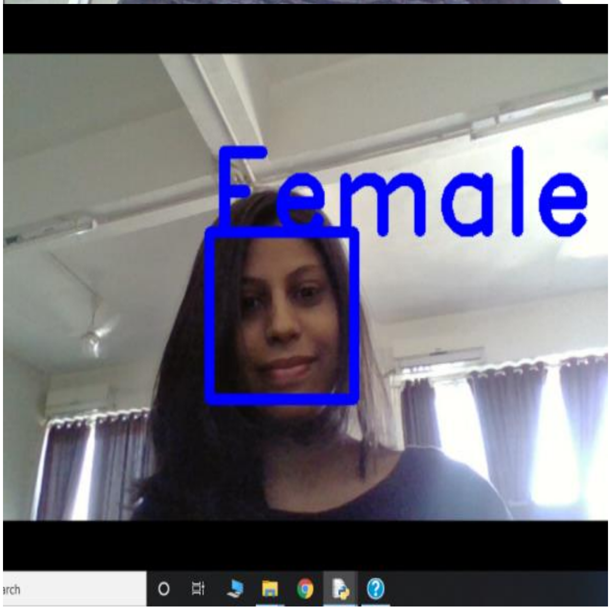
Because we would like to maximise the log likelihood of the proper class, the term we would like to reduce is that the negative log likelihood.

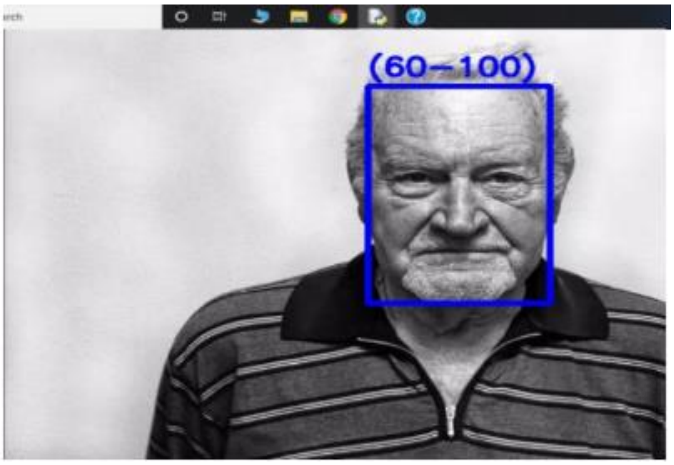
$$L_i = -log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

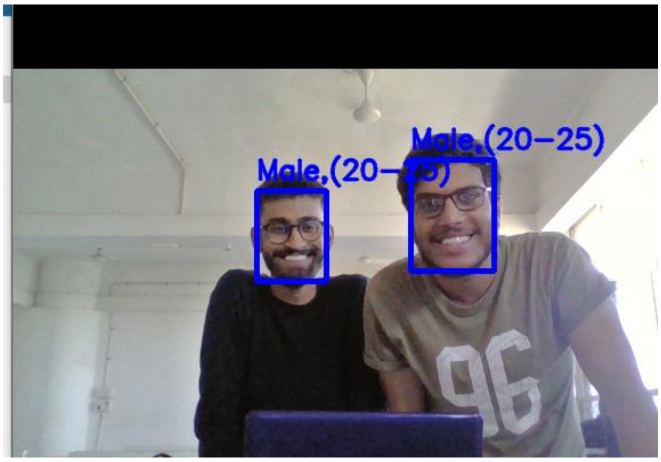
IV. RESULTS

Following are the results that we obtained after testing the real time images on our algorithm along with the accuracy.









The overall accuracy of the results is 89. 5.

V. CONCLUSION

The paper explains significant approach to gender recognition problem along with different approach for age approximation. This study was the trial of this idea. Although the results aren't perfect, they're promising to future studies. Our next step is to make a higher Haar cascade and use this method for other multiclass problems. Training the haar cascade classifier with much more data can surely improve the accuracy of the classifier. The easy availability of big image collections provides modern machine learning based systems with effectively endless training data, though this data isn't always suitably labeled for supervised learning. The system was programmed in python language. Both real time and static face detection was carried out. Taking example from the related problem of face recognition we explore how well deep CNN perform on these tasks using Internet data. We offer results with a lean deep-learning architecture designed to avoid over fitting because of the limitation of limited labeled data. The most difficult portion of this project was fitting the training infrastructure to properly divide the info into folds, train each classifier, cross-validate, and mix the resulting classifiers into a test-ready classifier. I foresee future directions building off of this work to incorporate using gender and age classification to help face recognition, improve experiences with photos on social media, and far more. Finally we hope that additional data training will improve the algorithm to provide better results.

REFERENCES

[1] J. Gou, L. Gao, P. Hou, and C. Hu, ―Gender recognition based on multiple scale textural feature,‖ presented at the 5th International Congress on Image and Signal Processing, Sichuan, China, October16-18, 2012.

[2] G. Levi and T. Hassner. Age and gender classification using convolutional neural networks. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops, June 2015.

[3] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimationof unfiltered faces. Trans. on Inform. Forensics andSecurity, 9(12), 2014S. Baluja and H. A Rowley.Boosting sex identification performance.Int. J. Comput. Vision, 71(1):111–119, 2007.

[4] W.-L. Chao, J.-Z.Liu, and J.-J. Ding. Facial age estimation based on label-sensitive learning and age-oriented regression. Pattern Recognition, 46(3):628–641, 2013

[5] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim. Age estimation using a hierarchical classifier based on global andlocal facial features. Pattern Recognition, 44(6):1262–1281,2011.

[6] C. Cortes and V. Vapnik. Support-vector networks. Machine learning, 20(3):273–297, 1995.E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. Trans. on Inform. Forensics and Security, 9(12), 2014.

[7] A.C.Gallagher and T. Chen. Understanding images of groups of people. In Proc. Conf. Comput. Vision Pattern Recognition, pages 256–263. IEEE, 2009

[8] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet:A neural network identifies sex from human faces. In NeuralInform. Process. Syst., pages 572–579, 1990d

[9] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang. Imagebasedhuman age estimation by manifold learning and locallyadjusted robust regression. Trans. Image Processing,17(7):1178–1188, 2008.

[10] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. A study on automatic age estimation using a large database. In Proc. Int Conf. Comput. Vision, pages 1986–1991.IEEE, 2009.Neural Inform. Process. Syst., pages 1097–1

[11] Y. H. Kwon and N. da Vitoria Lobo.Age classification from facial images. In Proc. Conf. Comput. Vision Pattern Recognition ,pages 762–767. IEEE, 1994

[12] C. Shan. Learning local binary patterns for gender classification on real-world face images. Pattern Recognition Letters,33(4):431–437, 2012.

[13] M. Toews and T. Arbel. Detection, localization, and sex classificationof faces from arbitrary viewpoints and under occlusion.Trans. Pattern Anal. Mach. Intell., 31(9):1567–1581,2009.

[14] Paul Viola and Michael Jones, Rapd object Detection using a Booscade of Simple Features. Accepted Conference on Computer Vision and Pattern recognition, 2001.

[15] Textural Features for Image Classification, (with K. Shanmugam and I. Dinstein), IEEE Transactions on Systems, Man, and Cybernectics, Vol. SMC-3, No. 6, November 1973, pp. 610-621.